Impacts of an early childhood mathematics and science intervention on

teaching practices and child outcomes

Jessica V. Whittaker [a], Mable B. Kinzie [a], Virginia Vitiello [a],

Jamie DeCoster [a], Christina Mulcahy [b], & Emily A. Barton [a]

[a] Center for Advanced Study of Teaching and Learning
Curry School of Education and Human Development
University of Virginia

[b] Marsico Institute for Early Learning and Literacy
Morgridge College of Education
University of Denver

Note: This article was published online March 6, 2020 (advance online publication) and is available at the following citation:

Whittaker, J. V., Kinzie, M. B., Vitiello, V., DeCoster, J., Mulcahy, C., & Barton, E. A. (2020). Impacts of an early childhood mathematics and science intervention on teaching practices and child outcomes. *Journal of Research on Educational Effectiveness*, *13*(2), 177-212. DOI: 10.1080/19345747.2019.1710884

The first author may be contacted at:
The Center for Advanced Study of Teaching and Learning
Curry School of Education
University of Virginia
405 Emmet St. South, #248
Charlottesville, VA 22904-4273

jwhittaker@virginia.edu
434-243-4905

Acknowledgements: The research reported here was supported by the U.S. Department of Education, Institute of Education Sciences through Grants R305A120631 and R305B140026. The contents in this publication are those of the authors and do not represent views or policies of the Institute of Education Sciences. We graciously acknowledge the many teachers, parents, children and research staff who made this study possible.

Abstract

This randomized controlled trial examined effects of the MyTeachingPartner-Math/Science intervention on the quality and quantity of teachers' mathematics and science instruction, and children's mathematics and science outcomes in 140 pre-kindergarten classrooms. Teachers participated in the intervention for two years with consecutive cohorts of children. Results from Year 1 are considered experimental, however due to high levels of attrition, results from Year 2 are considered quasi-experimental. Across both years, intervention teachers exhibited higher quality and quantity of instruction. In Year 1, there were no significant effects of the intervention on children's outcomes. In Year 2, children in intervention classrooms made greater gains in teachers' ratings of mathematics and science skills and performed better on a spring assessment of science skills. These results have implications for designing and evaluating professional development aimed at supporting children's mathematics and science knowledge and skills.

Impacts of an early childhood mathematics and science intervention on teaching practices and child outcomes

The foundational mathematics and science skills that make it possible for children to fully operate in the world around them are first developed in early childhood (National Association for the Education of Young Children [NAEYC] & National Council of Teachers of Mathematics [NCTM], 2002; National Research Council [NRC], 1996). However, for children growing up in poverty or experiencing other social disadvantages, an “opportunity gap” (Gutiérrez, 2008) may result from inadequate exposure to early experiences and high-quality interactions that support children’s learning.

Unfortunately, opportunities for learning are often missed in the early childhood classroom. Mathematics and science are covered in cursory ways (Nayfeld, Brenneman, & Gelman, 2011) or in discrete units lacking linkage across broad themes (NAEYC & NCTM, 2002; NRC, 2005). These limitations may be partially a function of teacher beliefs that young children are not developmentally ready to learn formal mathematics and science concepts (see the review of Piagetian child development theory by Gelman & Brenneman, 2004), but research has not supported these beliefs. Instead, research suggests that preschoolers demonstrate capacity for complex and abstract thought (National Mathematics Advisory Panel, 2008; NRC, 2000a) and that young children’s learning in these domains can be carefully scaffolded (NRC, 2007b).

Given these misconceptions, classroom curriculum is likely to be important because it provides clear guidelines for learning objectives and selecting/implementing activities that meet children’s needs. Indeed, high quality curricula have been found to support teachers in offering experiences to encourage children’s learning (Clements & Sarama, 2007a; 2008; French, 2004;

Greenfield, Jirout et al., 2009; Kinzie et al., 2014; Phillips et al., 2017; Presser, Clements, Ginsburg, & Ertle, 2015), with gains found for children's development of mathematics, language, literacy, and approaches to learning. At the same time, however, large scale national studies employing standardized observations show that, even when offered validated curricula, pre-k teachers often struggle to implement them with high quality and fidelity. This is likely due to teachers' lack of subject-area content knowledge and confidence (Pianta et al., 2005), a problem of particular importance in mathematics and science.

In response to these challenges, the standards-based MyTeachingPartner-Math/Science (MTP-M/S) curricula and teacher support system was developed for pre-k teachers through a research-based curriculum development process (Kinzie, Whittaker, McGuire, Lee, & Kilday, 2015). What makes the MTP-M/S system unique from most other curricula is that teacher professional development (PD) is embedded within the curricula as well as supported through online teaching demonstrations and a limited number of face-to-face teacher workshops.

In this manuscript, we report the results of a two-year trial in which the MTP-M/S curricula and teacher support system was compared to Business as Usual (BaU) with 140 participating teachers across a range of early childhood settings in two large metropolitan areas of the United States. Teachers participated in the intervention across two years with two consecutive cohorts of students. Year 1 used a randomized controlled design where teachers were assigned to condition. In Year 2, due to high levels of attrition and potential threats to the validity of the random assignment, the study is considered quasi-experimental. In both years, we determined associations between MTP-M/S and teachers' practice (including quality of interactions, quality of mathematics and science instruction, and dose of mathematics and science activities) and children's mathematics and science learning. We also consider the extent

to which associations between the intervention and children's outcomes are mediated through improvements in teacher practices. To situate the research reported here, we begin with an overview of children's early development of mathematics and science knowledge and skills. We describe our theoretical framework which posits that high-quality PD, including early childhood curricula and supports for implementation, have direct positive impact on teachers' interactions and instructional practices and child outcomes, and that there may be indirect effects on child outcomes through teacher practice. Then we go on to describe the design of the MTP-M/S intervention, briefly reviewing its research-based development. We close this background section with a summary of a previous field trial and the results for children's learning and teachers' practice.

Supporting Children's Development of Mathematics/Science Knowledge and Skills

Research over the past several decades indicates that young children (ages 0 to 5) develop early informal everyday mathematics skills that are both broad and complex (Ginsburg, Lee, & Boyd, 2008) and at times, sophisticated (Zur & Gelman, 2004), prior to any formal education. This development typically includes ideas involving basic number sense and operations (Baroody, Lai, & Mix, 2006; Bryant, 1995; Clements & Sarama, 2007b), counting (Baroody, 1992; Frye, Braisby, Lowe, Maroudas, & Nicholls, 1989; Gelman & Gallistel, 1978; Stock, Desoete, & Roeyers, 2009; Wynn, 1990), and geometric thinking (e.g., size, shape, location, and patterns; Clements, 2004a; Clements, Swaminathan, Hannibal, & Sarama, 1999). Basic problem solving skills are developed during this time, along with an understanding of simple calculation concepts (Levine, Jordan, & Huttenlocher, 1992).

In a similar fashion, research demonstrates that young children can come to understand scientific concepts such as the life cycle, growth and change, and distinctions between animate

and inanimate objects (Backscheider, Shatz, & Gelman, 1993; Inagaki & Hatano, 1996; Springer & Keil, 1991), and can reason scientifically. For example, children can infer how misleading evidence can encourage formation of a false belief (Ruffman, Olson, Ash, & Keen, 1993). By age six, children can differentiate between hypotheses and evidence (Ruffman, Perner, Olson, & Doherty, 1993; Sodian, Zaitchik, & Carey, 1991), which is earlier than prior research had suggested (Kuhn, 1989; Piaget & Inhelder, 1969).

As these findings suggest, functional expertise involves a grasp of factual knowledge and skills as well as the understandings that make this knowledge and skill "usable" (Bowman et al., 2001, p. 185). Critical thinking skills make these understandings possible. Central is the ability to analyze, make inferences, reason, and make decisions to solve problems (Lai, 2011). Although some assert that these skills are domain-specific, other researchers note that there is a core set of skills, applicable across domains and important for school success: discovering, assessing, revising, and communicating knowledge (Klahr, Zimmerman, & Jirout, 2011), capacities foundational to inquiry (Minner, Levy, & Century, 2010). Our research team selected mathematics and science as domains of curricular development for all of the reasons we have identified: The critical thinking skills developed by engaging in inquiry in these areas, the ability to integrate multiple content areas (e.g., language, literacy, etc.) into mathematics and science activities, and the importance of early mathematics and science knowledge and skills in supporting children's later academic development.

In response to the needs and opportunities for early childhood learning, an increasing number of mathematics curricula and associated PD efforts have been designed, aimed at improving the effectiveness of mathematics and science instruction (e.g., Clements, Sarama, Spitler, Lange, & Wolfe, 2011; Presser et al., 2015; Starkey, Klein, & Wakeley, 2004). Despite

the large investments in classroom curricula and PD, evaluations of these products are infrequent (NRC, 2009). A review by the National Center on Quality Teaching and Learning (NCQTL, 2015) indicated that there were nine commercially available mathematics curricula for pre-kindergarten, but that only three had been rigorously evaluated: Big Math for Little Kids (strength of effects on child outcomes rated as 3 out of 4; Greenes, Ginsburg, & Balfanz, 2004), Building Blocks (strength of effects rated 4 out of 4; Clements & Sarama, 2007c) and Pre-K Mathematics (strength of effects rated 4 out of 4; Klein, Starkey, & Ramirez, 2002).

Like with mathematics, science curricula can serve as "cognitive tools that are situated in teachers' practice" (Davis & Krajcik, 2005, p. 3), thus providing effective scaffolds for teachers and students (Zucker, Tinker, Staudt, Mansfield & Metcalf, 2008). However, the development and evaluation of pre-kindergarten science curricula is even more infrequent than in the domain of mathematics. Some notable exceptions are ScienceStart! (French, 2004), PreSchool Pathways to Science (PrePS; Gelman & Brenneman, 2004), the Young Scientist Series (Chalufour & Worth, 2003; 2004; 2005), Early Childhood Hands-On Science (ECHOS; Brown & Greenfield, 2010; Greenfield, Jirout, et al., 2009), and Foundations of Scientific Literacy (Gropen, Kook, Hoisington, & Clark-Chiarelli, 2017). Each clearly addresses learning objectives in alignment with the national standards including an emphasis on science inquiry skills such as prediction, observation, and recording and analysis of observations, as well as relevant vocabulary. However, there is variability in the degree to which these curricula have been rigorously evaluated to determine whether they are associated with gains in children's science achievement. Additionally, not all of these curricula include opportunities for teachers to receive PD and training that could support their implementation.

Characteristics of Effective Professional Development

The benefits associated with well-designed curricula are the greatest when teachers receive not only strong curricula, but the support needed to help them implement them well and promote children's skill development through their classroom interactions (e.g., Pianta, Mashburn, Downer, Hamre, & Justice, 2008; Wasik, Bond, & Hindman, 2006). Expert recommendations for teacher development in mathematics suggests that key components include a focus on specific, well-articulated objectives and teaching practices (including high quality interactions), opportunities for active learning and discussion of strategies, and sustained participation over time (Garet, Porter, Desimone, Birman, & Yoon, 2001; NRC, 2006; Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010).

Most often, effective curricular packages offer embedded teacher supports aimed at developing teachers' concept knowledge as well as their pedagogical expertise. Embedding PD support in the context of instructional materials can encourage transfer of related knowledge and skills to teaching practice. These embedded supports take the form of curricular highlights on relevant concept knowledge (and how a student at that level might best come to understand it and potentially misunderstand it), the developmental trajectories of children's knowledge and skill development and child assessments to aid teachers in recognizing where students are in their development, best instructional practices, and recommendations for differentiating instruction to meet their students' needs (e.g., Clements & Sarama, 2007c; Gelman & Brenneman, 2004; Greenfield, Jirout, et al., 2009; Kinzie et al., 2015).

On-going workshops and coaching represent another common format for teacher PD. Yoon, Duncan, Lee, Scarloss, and Shapley (2007) found that an average of 49 hours of PD enabled teachers to improve their students' achievement; this is roughly the amount employed to

good effect by Clements and Sarama (2008) in their implementation of the *Building Blocks* curriculum. The NRC (2007a) has recommended delivery of effective PD via the Internet, to help increase scalability and accessibility, and as a result, be more likely to make a detectible difference in the practice of a large number of teachers. In addition, the media capabilities provided by Internet delivery make video demonstrations of optimal teaching practice possible and 24/7 availability ensures that online supports can more readily be utilized by busy teachers while minimizing out-of-classroom and travel time. The cost of PD may also be decreased with online delivery: the publishers of *My Math*, a popular pre-k mathematics curriculum, offer training webinars for less than half of the cost of in person training (NCQTL, 2015).

Finally, a factor thought to be especially important in supporting teachers' curricular implementation is multiple years' experience with the curriculum. The positive effects reported in some studies of early childhood mathematics curricula involve teachers' second year of implementation (e.g., Clements et al., 2011; Preschool Curriculum Evaluation Research Consortium [PCER; in which 80% of teachers were in their second year of curricular implementation], 2008). A second year of implementation for the Enhanced Reading Opportunities study resulted in better implementation fidelity in the 34 participating high schools: 26 schools were observed to be well-aligned in year two compared to 16 in year one, and the number of poorly aligned schools decreased from 10 in year one to only one in year two (both years resulted in significant gains in students' reading comprehension; Kemple et al., 2008). Accordingly, an ongoing program of curricular implementation can be key to supporting teachers' practice.

In light of the above literature, we designed the MTP-M/S curricula to capitalize on the potential for early learning in mathematics and science, and to incorporate multiple forms of

teacher PD to support teachers' implementation of the curricula. Next, we describe our theory of change, the design of curricular and PD resources, and their evaluation to-date (see also the descriptions of the curricula, and the research that led to it, in Kinzie et al., 2015).

The MTP-M/S Intervention

Theory of change. Our theory of change posits that PD that includes high-quality curricula and support for high-quality implementation of the curricula can lead to changes in both child outcomes and classroom practice, including teacher-child interactions, the quality of mathematics and science instruction, and the amount of instruction. We also hypothesize that PD may affect child outcomes indirectly through changes in teachers' practice (see Figure 1 for our logic model). In other words, effective PD leads teachers to improve their classroom practice and those changes lead to greater child learning.

As outlined above, there is a growing research base suggesting that mathematics and science PD that includes a curricular component and support for implementation can have positive impacts on the quality of teacher-child interactions, mathematics and science instruction, and the amount of instruction taking place in the classroom (e.g., Gropen et al., 2017; Piasta, Logan, Pelatti, Capps, & Petrill, 2015; Sarama, Clements, Wolfe, & Spitler, 2016), and in some cases, direct effects on child outcomes (Clements et al., 2011; Klein, Starkey, Clements, Sarama, & Iyer, 2008; Presser et al., 2015).

In our design of the MTP-M/S curricula and teacher supports, we have drawn on the substantial scholarship in early mathematics and science teaching and learning to include support for high-quality teacher-child interactions and mathematics and science instruction. MTP-M/S curricular activities and associated PD were designed to:

- Build on children's natural curiosity and draw upon children's everyday experience

(Conezio & French, 2002; Harlan & Rivkin, 2004; Yoon & Onchwari, 2006), challenging students to find the mathematics (e.g., mathematize) and science that are inherent in these experiences, and to develop related understandings (Sarama & Clements, 2003).

- Help students develop a continuum of understanding, ranging from construction of instrumental understandings (e.g., rules, procedures, and definitions) to the evolution of relational understandings as these concepts are applied to solve problems (Skemp, 1976); such that the connections between them and children's existing understandings are strengthened (Carpenter & Lehrer, 1999; NRC, 2000b; Sousa, 2008).
- Engage children in inquiry activities, by predicting, observing, experimenting and communicating (Gelman & Brenneman, 2004; NRC, 2006).
- Focus on mathematical and scientific language development, by promoting teachers' "math [and science] talk" (Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006) and encouraging students to use language to express and justify their thinking (Ginsburg et al., 2008).

There is evidence that these proximal classroom practices, in addition to high dosages of mathematics and science content, are associated with growth in children's mathematics and science outcomes (e.g., Klein et al., 2008; Piasta et al., 2015). There is a more limited research base that has explored whether the provision of PD indirectly impacts children's knowledge and skills through an impact on classroom practice, especially in the areas of mathematics and science, although there are a few studies that have suggested that teachers' practice might be the mechanism through which PD may impact child outcomes (e.g., Piasta et al., 2015; Sarama et al., 2016).

Curricular design. Given the opportunity gap evident for children from poor or disadvantaged families (Gutiérrez, 2008), we designed MTP-M/S to serve pre-kindergarten students at-risk of early school failure. Research indicates the value of both "student-centered" and "teacher-directed" approaches (National Mathematics Advisory Panel, 2008; Snow, 2011), so in the curricular design we followed the advice of Justice and Pullen (2003) and aimed for a balance between them. Drawing from the theoretical frames of structured inquiry (NRC, 2006), situated cognition (Brown, Collins & Duguid, 1989), and cognitive development (e.g., Ginsburg & Golbeck, 2004), we incorporated meaningful student-centered explorations relevant to children's everyday lives, balanced with explicit, teacher-guided, exposures to key mathematics and science concepts and skills.

Structured inquiry is an intermediate form of inquiry methods (NRC, 2006), in which teachers pose questions and suggest procedures, but children actively observe, predict, collect, analyze, and communicate their processes and results. In the MTP-M/S curricular design, students engage in inquiries of increasing sophistication across the year, with authentic events providing the context for collaborative knowledge construction and skill acquisition. For instance, instead of a static month-long "unit" on plants, students' explorations are linked as much as possible to the seasons and what is happening outside. In the fall, students explore the environment near their schools and collect seeds as they ripen. In winter, students sort and describe seeds and consider their use as a food source. In early spring, students examine seeds (and the "baby plants" that emerge) before planting the sprouts. In later spring, students experiment to study the effects of light and water on the growth of these plants, just as plants begin to flourish outside. Children's cognition is expressly cultivated within these situations. Building on the work of Ginsburg and Golbeck (2004), the activity designs emphasize thinking

(reflective, as a function of inquiry, and also problem solving) and the modeling and eliciting of mathematical and scientific language to express and give form to that thinking. Children's texts introduce key concepts and also helps to develop students' literacy.

The MTP-M/S mathematics and science curricula are aligned with national and state standards. For mathematics, the curriculum builds upon the NCTM curriculum focal points for pre-k (2006), Clements' (2004b) developmental trajectories for grades P-2, and a review of state standards. The mathematics domains covered include: number sense, operations, geometry, and measurement. Within each domain, the "big ideas" (the most important concepts and skills; Ginsburg & Ertle, 2008; Greenes, Ginsburg, & Balfanz, 2004) are addressed in a developmental progression across the year, informed by reviews of the research (Clements & Sarama, 2004; Kinzie et al., 2015). For example, during an inquiry related to measurement, children develop key conceptual understandings about the relative weight of objects and how these relative weights can be accurately determined. They begin by examining a golf ball and a ping-pong ball and predicting which is heavier. Next, they test their predictions through use of a balance and compare the results to their predictions. Finally, they extend their understandings to prediction, testing, and reflection of a variety of objects across the day.

In science, the Benchmarks from the American Association for the Advancement of Science (AAAS, 1993) and the National Science Education Standards (NRC, 1996) articulate trajectories for grades K-2 and K-4, respectively. A review of state pre-k standards was used to validate these targets and refine the curricular focus for pre-k. The MTP-M/S Science curriculum addresses three domains of science: life science, earth science, and physical science, with inquiry-based activities to meet instructional objectives aligned with state and national standards. The activities guide children in construction of conceptual understandings across

these broad domains. Activities provide opportunities for students to apply inquiry skills to predict, observe, analyze, and describe their findings, and to use simple science tools.

The MTP-M/S curricular packages include two mathematics and two science activities for each of the 33 weeks of the school year for a total of 132 activities, all of which offer within-activity and online implementation supports. Activities were designed to be 15-20 minutes in length and were in either whole or small group format. For small group activities, teachers were asked to conduct the activity multiple times so that all students would have the opportunity to participate (see Figure 2 for a sample activity from the MTP-M/S curricula). In addition, one mathematics or science "center time" investigation is provided each week, extending one of the week's curricular activities. Teachers were asked to introduce the center time activity at the beginning of the week, and then allow students to explore them independently. We provided suggestions for how teachers could support students as they were exploring the center. Monthly "Do It at Home" newsletters offer activity suggestions that parents/caregivers can engage in with their children to support knowledge and skill development in alignment with curricular objectives.

Hybrid teacher professional development system. To promote teachers' related knowledge and skill development, the hybrid teacher support system is informed by the situated cognition theory, with supports anchored to teachers' authentic tasks and the encouragement of teachers' reflection on practice (Brown et al., 1989; Cognition and Technology Group at Vanderbilt, 1990). These supports are delivered "within-activity," online, and via workshops. The goal with this hybrid approach was to provide supports in multiple media formats, to reduce workshop time requirements, and to maximize support available to teachers on a "just-in-time" basis.

Within activity supports. Acknowledging observed limitations in preparation of early childhood teachers, the NRC (2005) called for curricula comprehensive enough for teachers with a range of preparation and background knowledge and experience to implement successfully. Extensive teacher instruction can require a substantial amount of time for teachers to process, so finding a balance was important in developing effective support components that were still "consumable" within the time teachers could devote. To achieve this aim during the iterative development process (Kinzie et al., 2015), educative curricula were developed, promoting teacher learning with the actual curricular materials (Forbes & Davis, 2010). All activity instructions were distilled to what was most important for teachers to know and children to experience. Activities were formatted with a common four-step inquiry model: Engage, Investigate, Discuss, and Extend. Within this focused frame, the following occur:

- Every activity is preceded by a brief teaching tip on best pedagogical practices, common ways children construct understandings, including misconceptions, or key mathematical or scientific concepts.
- Varied, open-ended reflection questions are integrated throughout activities, supporting teacher/child reflection on what they observe and encouraging discussion about what observations might mean.
- When asking children to explain their thinking, teachers are provided with specific mathematical and scientific language to model for and to elicit from them.
- Adaptations are offered to better differentiate the learning activities to accommodate children of differing abilities while maintaining support for high quality instructional interactions and intellectual challenge that adhere to the goals of the curricula.

- Extensions enabling application of children's emergent knowledge and skills to a range of other activities across the day are suggested.

***Online supports*.** Online modalities can offer high value in support of teaching practice by allowing for enhanced media in support materials such as authentic enactments of curricular implementation (Davis & Krajcik, 2005). As part of the MTP-M/S PD support system, teachers are encouraged to review video demonstrations of high quality, high fidelity implementations of the MTP-M/S activities. Such exemplars have been found to be more effective than textual descriptions in encouraging pre-service teachers to describe specific applications of targeted teaching skills (Moreno & Ortegano-Layne, 2008). These include over 130, three-to-five minute videos created from recordings of teachers from our target population, with at least one demonstration provided for every curricular activity. We tracked teachers' use of web-based resources through a server that automatically recorded information about the number of logins and the length of time spent on the website. Logins to the MTP-M/S website ranged from one to 110 times in Year 1 (M = 12.03, SD = 17.01) and one to 68 times in Year 2 (M = 6.13, SD = 10.53). Total minutes spent on the website ranged from 1.53 to 2,211.42 minutes in Year 1 (M = 189.63, SD = 355.74) and 0.13 to 1,899.58 minutes in Year 2 (M = 98.01, SD = 267.61).

***Workshop-based supports*.** Teachers attend an introductory workshop in the summer where they are oriented to the curricula, materials, and PD. During the academic year, teachers' processing of the within-activity and online supports is deepened through a series of 2.5-hour PD workshops (five across their first year of implementation, and three in their second). The workshops are designed to enhance teachers' use of the online supports, encourage self-reflection and peer discussion of their own teaching practice, and support their confidence in facilitating high-quality learning experiences that are in alignment with the curricula. In Year 1,

teachers attended an average of 2.84 workshops (SD = 1.96), and in Year 2, an average of 0.91 workshops (SD = 1.20).

Previous Findings

A small, randomized trial compared the impacts of the MTP-M/S curricula and teacher support system (the Plus condition) to the MTP-M/S curricula alone (Basic) and to a BaU control group (Kinzie et al., 2014; Whittaker, Kinzie, Williford, & DeCoster, 2016). In analyses controlling for students' performance on fall assessments as well as demographic characteristics, students in the Plus group (g = 0.61) and the Basic group (g = 0.52) experienced significantly greater gains in geometry and measurement than students in the BaU group. Plus students also out-performed those in both the Basic (g = 0.35) and BaU (g = 0.47) groups on a measure of number sense and place value. There were no significant treatment effects on children's gains in life science (g's between 0.03 and 0.37) or earth and physical science (g's between 0.01 and 0.26). Teacher outcomes also showed positive effects for the MTP-M/S intervention, with teachers in the Basic and Plus conditions showing higher levels of Instructional Support (*effect sizes (ES)* = 0.58 and 0.60, respectively) and Facilitation of Mathematical and Scientific Thinking (ES = 0.38 and 0.44, respectively) compared to teachers in the BaU condition. Teachers in the Basic condition also showed higher levels of Emotional Support compared to teachers in the BaU condition (ES = 0.27).

Current Study & Research Questions

In the current study, we extend previous research to explore the effects of the MTP-M/S curricula and teacher support system on teachers' practice and children's mathematics and science outcomes. We used a randomized control trial in which teachers were assigned to the

MTP-M/S treatment or BaU condition to explore our logic model and answer the following research questions:

1. Is MTP-M/S associated with improvements in teacher practice (i.e., teacher-child interactions, mathematics teaching practice, science teaching practice, dosage of instruction) during the preschool year relative to the comparison group?
2. Is exposure to implementation of MTP-M/S associated with improvements in children's mathematics and science outcomes during preschool relative to the comparison group?
3. Are the associations between MTP-M/S and children's mathematics and science outcomes during preschool mediated through improved teaching practice?

Method

Participants

Participants in the study were recruited from two sites. Both were in large urban areas, one in the mid-west and one in the southeast. Teachers from both sites participated for two years with two consecutive cohorts of children. Site 1 teachers participated from 2013 – 2015, and were in a variety of settings including public, private, and Head Start classrooms. Site 2 teachers participated from 2014 – 2016, and were all in public pre-k classrooms. Teachers were eligible to participate if they had access to an Internet-connected computer and had at least 12 children enrolled, of whom 75% or more were eligible for kindergarten the following year, were English speakers, and were typically-developing. One hundred fifty-six teachers and their classrooms were initially recruited into the study (116 from Site 1, 40 from Site 2). Sixteen classrooms (17 teachers) left the study before child recruitment or child assessment began (three classrooms were not eligible based on child enrollment, one program closed, six teachers withdrew from the study, nine teachers left the center or classroom, one of these nine was replaced and then dropped

before data collection began). The analysis sample therefore includes 140 classrooms who were recruited into the study, met eligibility requirements, and for whom child data are available. There was significant attrition over the study period, with 71 classrooms leaving the study. The most common reason was teachers leaving the center or classroom (n = 64), followed by teachers who withdrew from the study (n = 7), and classrooms that were no longer eligible due to changes in enrollment (n = 4). In the first year only, replacement teachers were recruited into the study when possible, in order to retain the classroom (n = 14); 10 of these teachers ultimately left their classroom or center and were therefore dropped from the study. In all, 69 classrooms completed two years of participation (32 control, 37 treatment; See Figure 3 for Consort Flow Diagram). Teacher demographics are presented in Table 1.

Parental consent was obtained for 3,625 children, from which approximately six children (M = 6.40, SD = 1.10, min = 2, max = 9) per classroom per year were randomly selected for fall and spring assessment (n = 1,266 across both study years). Children were eligible to participate if they were 4 years old and eligible to attend kindergarten the following year, had no Individualized Education Plan (excluding education plans for speech), and were English speaking. Two hundred thirty-five children left the study due to teacher attrition, a determination that the child was no longer eligible to participate (1), the child leaving the classroom or school (20), or for unknown reasons (113). When possible, additional children were selected to replace those who left the study, resulting in the replacement of 105 children during the study. The total child sample, comprised of all children who were assessed at either fall or spring of either year, included 1,371 children (720 treatment, 651 control). Child and family demographic characteristics are presented in Table 1.

Procedures

Recruitment and random assignment. For Site 1, researchers identified a list of childcare centers to which we sent flyers describing the project and followed up by contacting centers individually. If center directors indicated that they were interested in participating, researchers held a recruitment meeting with teachers to describe the project in more detail and obtain teachers' consent. For Site 2, researchers recruited a public school-based preschool program. Project information was distributed to teachers by center directors. Teachers who opted to participate returned consent forms to the research team. Once we received a consent form, a teacher was considered to be enrolled in the study.

After enrollment, stratified random assignment was used to place teachers in the MTP-M/S treatment or BaU control condition. We stratified by the number of participating teachers per center and type of center (i.e., Head Start, public pre-k, private pre-k). For centers with two or four study classrooms, half were randomly assigned to each condition. For centers with three classrooms, half were randomly selected to have two randomly assigned treatment classrooms and one control classroom, and half were selected to have two randomly assigned controls and one treatment. Programs with a single classroom were grouped according to center type, and half of each group was randomly assigned to each condition.

At the beginning of the school year, participating teachers sent all parents or guardians of their students a consent form and short family demographic survey. Eighty-eight percent of parents consented to allow their children to participate in the study (3,625 out of 4,075 parents).

Orientation and materials. At both sites, both treatment and control teachers attended an orientation describing basic procedures for the study. Treatment teachers then attended a workshop that included an overview of the curricula and implementation supports. The treatment group received the MTP-M/S curricula, teaching materials needed to implement the

activities, and access to online supports as described above. BaU teachers were asked to continue teaching mathematics and science as they normally would. To control for any Hawthorne effects (McCambridge, Witton, & Elbourne, 2014), they were invited to attend the same number of PD workshops as MTP-M/S teachers, although the topics focused on supporting children's social-emotional development instead of their mathematics and science learning.

Data Collection

Teacher and family surveys. Participating teachers and parents completed demographic questionnaires in the fall. Teachers reported their age, experience, and education. Parents reported family income, household composition, parent education, and children's age, gender, and race.

Video collection. Teachers were provided with digital cameras, SD cards, and postage-paid mailers and were asked to film all of the mathematics and science lessons they taught each month. Treatment teachers were asked to implement and record all MTP-M/S activities (two mathematics and two science per week across 33 weeks of the school-year). They were asked to record one activity per tape, and to submit 12 - 16 tapes per month, corresponding to the number of curricular activities provided, for a total of 132 tapes across the year (for the months of November, December, and April, we only provided three weeks' worth of curricular activities to account for vacation time). Treatment teachers enclosed a lesson log with their videos that noted each lesson they taught and whether the lesson had been recorded. The research team checked each treatment SD card against the teacher's log before noting that the video was received and what content it contained. Treatment teachers submitted an average of 73 tapes in Year 1 (SD = 43) and 70 tapes in Year 2 (SD = 42). Control teachers were asked to film all mathematics and science lessons, plus other activities if necessary to reach 12 - 16 videos per month. This was to

ensure that teachers in both conditions submitted the same number of videos per month. Control teachers also completed a log noting the content of each recorded lesson. For these teachers, the research team screened all videos to confirm the content. Control teachers submitted an average of 44 tapes in Year 1 (*SD* = 30) and 44 tapes in Year 2 (*SD* = 32).

Video observations. Mathematics and science teaching quality, as well as teacher-child interactions, were coded from teachers' videos. Out of a possible 12-16 videos submitted per month (number of possible videos differed depending on the month, as described above), we selected two tapes for each teacher for coding, for a total of 12 tapes per year. We selected videos for coding in an effort to obtain an adequate sample of teachers' practice across the year (September, October, November, February, March, April), across domains (mathematics and science), and across activity settings (whole vs. small group). Teachers were randomly assigned to one of two selection schemes: whole group mathematics/small group science or small group mathematics/whole group science, which alternated each month. If a teacher did not submit a video that fit their selection for a given month, we were flexible in replacing videos that did not fit the group size criteria (e.g., substituting a whole group activity for a small group activity), but did not substitute videos that failed to meet the domain criteria (e.g., we did not replace a mathematics activity with a science activity). On average, 8.10 (*SD* = 4.10) tapes were coded per teacher in Year 1 and 8.90 (*SD* = 3.90) in Year 2 (*min-max* = 2 – 15 both years). Video coding teams were trained to reliability on the measures described below and met weekly for calibration sessions, which involved coding and discussing master-coded videos. Videos, including those which were double coded to assess reliability, were randomly assigned to coders.

Mathematics teaching quality was rated using the Classroom Observation of Early Mathematics Teaching (COEMET; Sarama & Clements, 2007), which includes five quality

items rated on a 5-point scale. The quality items assess teachers' presentation of content, instructional scaffolding, and responses to children as they relate specifically to mathematics instruction. The COEMET demonstrated sensitivity to change in instructional quality in a previous randomized trial of a mathematics intervention (Hofer, Lipsey, Dong, & Farran, 2013). We double coded 100% of all videos and inter-rater reliability was acceptable (*ICC* average measures = .78). Internal consistency (α = .72) was also acceptable.

Science teaching quality was rated using the Preschool Science Observation Measure (PSOM; Vitiello, Whittaker, Kinzie, Mulcahy, & Helferstay, 2018) developed by the study authors. The PSOM captures teachers' efforts to engage children in science content, scaffold knowledge and vocabulary, and connect information to existing knowledge. Seven quality items were rated on a 7-point scale with detailed descriptions of low, mid, and high quality science instruction. Scores on the seven items were averaged together to create a total science quality score with strong internal consistency (α = .87). Preliminary analyses indicated that PSOM science quality was significantly correlated with the CLASS total score (r = .57, p < .001), providing evidence of convergent validity. The PSOM total score was also significantly associated with higher scores on the Preschool Assessment of Science (Gropen et al., 2017) direct assessment in Year 2 of the study. Inter-rater reliability analyses were conducted on 20% of the ratings and showed acceptable reliability (Cicchetti et al., 2006; *ICC* = .75).

Global quality was assessed using the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). The CLASS captures teacher-child interactions in three broad domains: Emotional Support, Classroom Organization, and Instructional Support. Ten dimensions are rated on a 7-point scale based on detailed descriptions of low, mid, and high quality interactions, which are averaged to create domain scores. The CLASS has been widely

used to assess preschool quality and has shown positive links to child outcomes (Mashburn et al., 2008). One hundred percent of videos were double coded and inter-rater reliability analyses indicated acceptable agreement (Cicchetti et al., 2006; Emotional Support, *ICC* = .73; Classroom Organization, *ICC* = .71; Instructional Support, *ICC* = .65).

Live Observations. The proportion of time spent teaching mathematics and science were coded during live classroom observations each spring. Data collectors completed the Classroom Snapshot (adapted from Ritchie, Howes, Kraft-Sayre, & Weiser, 2001), which allowed observers to track teachers' use of class time and the content of instruction and activities. Observers continuously tracked classroom activity settings (e.g., whole group instruction, small group instruction, individual time, book reading) and content (e.g. literacy, mathematics, social studies) using an iPad timing application during 15-minute cycles across the morning, lasting an average of two hours per observation (M = 2.00, SD = 0.77). Coders visited each classroom for an average of 3.30 days in Year 1 and Year 2 (SD = 0.81 and 0.64, respectively) and proportions were averaged across the observations. We observed across three days in order to try to get an accurate estimate of the amount of time that teachers spent engaging in mathematics and science instruction. In the current study, inter-rater agreement was good (Cicchetti et al., 2006; proportion of mathematics, *ICC* = .66; proportion of science, *ICC* = .76).

Child assessments. All child assessments were conducted in the fall and spring of each study year, except for the Preschool Assessment of Science (PAS; Gropen et al., 2017), which was only collected during the spring. Data collectors completed a two-day training to learn the measures prior to assessing children. Children were assessed in a quiet space, outside of the classroom when possible. Assessments were divided into two sessions on the same day to prevent child fatigue.

Children's mathematics and science achievement were assessed using teacher rating scales and direct assessments. Teachers completed the Academic Rating Scale (ARS) Math and Science for each child, adapted from the Early Childhood Longitudinal Study-Kindergarten Cohort's (ECLS-K) Academic Rating Scale (National Center for Education Statistics, 1998, 2011). Teachers used a 5-point scale (1 = *Not Yet* to 5 = *Proficient*) to rate children's mathematics and science skills. For mathematics, the original ARS had seven items; the research team developed five additional items to provide more complete coverage of mathematical skills children are developing in pre-kindergarten ($\alpha = .97$). For science, the original ARS included three items; the research team developed seven additional items reflecting a wider range of science conceptual understandings and skills ($\alpha = .97$).

Mathematics skills were directly assessed using an adapted version of the Short Tools for Early Assessment in Mathematics (MTP-STEAM) and an assessment of Number Sense and Place Value (NSPV; Kinzie et al., 2014). The STEAM, a short form of the commercially available TEAM assessment, includes items assessing a range of basic mathematics skills (counting, sequencing, number recognition; Weiland et al., 2012). It has been validated for use in pre-k and has shown sensitivity to individual child differences (Weiland et al., 2012). For the current project, additional items related to measurement (five items) and shapes (15 items) were added for a total of 41 items. The MTP-STEAM showed strong internal consistency ($\alpha = .82$).

The NSPV is a dynamic assessment of children's counting and number recognition, and understanding of place value (Kinzie et al., 2014). It consists of nine multi-part items that use progressive scaffolding to determine the level of support the child needs in order to demonstrate the target skill, with lower points awarded to children who require greater scaffolding. Children can score a total of 41 points. The NSPV showed adequate internal consistency ($\alpha = .67$).

Science skills were directly assessed using the LENS on Science (LENS; Greenfield, Dominguez, Greenberg, Fuccillo, & Maier, 2011) and the Preschool Assessment of Science (PAS; Gropen et al., 2017). The LENS is an adaptive IRT-based instrument that assesses children's knowledge of life science, earth and space science, physical and energy science, as well as science process skills. The LENS selects items for administration based on identifying items that have a difficulty parameter value that are closest to the estimate of a child's ability level. IRT analyses have demonstrated that LENS scores have high item reliability (.98) and person reliability (.93; Greenfield et al., 2011). The LENS is significantly and positively associated with direct assessments of vocabulary, mathematics, listening comprehension, and alphabet knowledge skills (Greenfield, Dominguez et al., 2009). In our study, LENS scores ranged from -1.07 to 3.00.

The PAS is a task-based assessment that requires children to predict, observe, reflect on, and explain scientific phenomena (Gropen et al., 2017). It includes two tasks, one related to floating and sinking and one related to understanding water level. For the floating and sinking task, children predicted whether three objects would sink or float, made observations by putting the objects in water, reflected on their predictions, and explained why one sank. For the water level task, children observed a tube partially filled with water, predicted what the water level would look like if the tube were placed at different angles, and explained why. Explanations were written down by data collectors and coded by members of the research team, blind to treatment condition, using a 5-point scale, with one corresponding to "no explanation" and five corresponding to "correct explanation using prior experience." The coding team consensus coded responses, first together and then separately, until agreement above 90% was consistently reached (~40 responses). They then coded the rest of the responses independently. There was a

total of 16 items; items were averaged to create a total score that showed marginal internal consistency ($\alpha = .41$).

Compensation. To thank them for their help, teachers in both conditions were compensated $175 for their assistance with data collection efforts and received $25 each month they submitted videos, for a total of up to $400 each year. Children were given stickers to thank them for their participation at the end of each assessment session.

Results

Descriptive and Preliminary Analyses

Descriptive and bivariate statistics were calculated without imputing or adjusting for missing values, as their purpose was to provide basic information on the collected variables. Table 1 presents the descriptive characteristics of the students and teachers enrolled in the study. Children were 4.50 ($SD = 0.32$) years of age. Approximately half of the children in the sample were White. Parents had an average of two years of post-high school education. Teachers were largely female (87%) and White (75%). Teachers had an average of three years of post-high school education and had substantial experience teaching pre-K. Table 2 presents the raw means and standard deviations of teacher practice and child outcome variables used in the study.

Attrition and Missingness. There was a moderate amount of classroom attrition by the end of Year 1 (19.2%), and a substantial amount of attrition by the end of Year 2 (55.8%). Table 3 provides the attrition rates for intervention and control classrooms at each time point in the study. We used chi-square tests to determine whether missingness was related to treatment condition at each of the four time points (fall Year 1, spring Year 1, fall Year 2, spring Year 2). None of these tests were significant (all ps > .38).

The What Works Clearinghouse (WWC) at the Institute for Education Science has produced a white paper describing how estimates of attrition bias can be obtained by examining the combination of differential and overall attrition (What Works Clearinghouse, 2018). The white paper identifies which combinations of differential and overall attrition meet WWC evidence standards, what combinations meet WWC evidence standards with reservation, and what combinations fail to meet WWC standards. The combination of differential and overall attrition in our sample at the Year 1 Fall and Spring time points meet WWC evidence standards. The attrition in our sample at the Year 2 Fall and Spring time points, however, do not allow us to meet WWC evidence standards without supplemental evidence of baseline equivalence. The high rates of attrition suggest that Year 2 should be considered a quasi-experimental design, given threats to the randomization (What Works Clearinghouse, 2017).

Baseline Equivalence. We conducted tests to examine whether our randomization resulted in statistically similar groups at baseline. Condition was not significantly related to child or teacher characteristics including child age, child gender, child race, parent education, family income, teacher age, or teacher experience (all ps > .30). We tested whether condition was related to child or teacher characteristics just for individuals who were in the sample at each of the time points. We never observed any significant relations of condition with child or teacher characteristics at any timepoint (all ps > .05).

We examined whether condition was significantly related to baseline assessments of any of the child outcome measures. There were no significant relations between condition and baseline assessments of child outcomes (ps > .10, respectively). Finally, we calculated effect sizes to determine the magnitude of the differences between the intervention and control groups on baseline child outcomes. The effect sizes for tests comparing the baseline equivalence of the

samples present at each of our four time points are presented in Table 4. Note that the effect sizes reported in this figure are always based on Fall Year 1 outcomes, but are calculated only using students that are still in our sample at the identified time point. There are only minimal differences between groups on all child outcomes in Year 1 and the ARS-Math, ARS-Science, and Lens in Year 2 (all d's < .10). When there is evidence of non-equivalence at baseline with effect sizes between 0.05 and 0.25, the WWC requires a statistical adjustment (WWC, 2018). We therefore controlled for baseline child outcomes in all subsequent analyses.

The magnitude of the effect size differences for the NSPV and MTP-STEAM in Year 2 were all greater than .30, failing to meet the WWC group design standards. Conclusions about treatment effect on such variables must be suspect because the groups are not equivalent. Although we still include these measures in our analyses below, we recognize the importance of this limitation and consider it further in the discussion section.

Bivariate Relations

Table 5 presents the correlations among the teacher practice variables for Year 1. Some notable results include the strong relation between mathematics quality and science quality (r = .67), the strong relations among the CLASS domains (rs = .54 to .77), and the strong relations of the CLASS Instructional Support domain with mathematics quality and science quality (r = .65 and .67, respectively). The proportion of time engaged in mathematics, proportion of time engaged in science, and number of videos submitted all have small or no relations with other measures of teacher practice.

Table 6 presents the correlations among the child outcome variables (aggregated to the classroom level) for the fall of Year 1. Since this time point occurs before substantial exposure to the intervention, these represent the pre-intervention correlations among these variables. All

of the measures are strongly related to each other, although teacher-rated mathematics and science ability are more strongly related than the others.

Table 7 presents the correlations of the fall Year 1 child outcomes with the Year 1 teacher practice outcomes. None of these relations are statistically significant (all p's > .01).

Research Question 1: Is MTP-M/S Associated with Improvements in Teacher Practice During the Preschool Year Relative to the Comparison Group?

Inferential models were run using classroom as the unit of analysis, with child-level variables aggregated to the classroom level. We attempted to examine all of the classroom outcomes simultaneously in a single path model, but this analysis did not converge. We therefore tested each outcome in its own, unique regression model. Year 1 and Year 2 outcomes were also modeled separately. The sample for Year 1 included the full 140 classroms (70 treatment, 70 BaU) enrolled at the start of the study. The sample for Year 2 included 75 teachers (40 treatment, 35 BaU) that were still enrolled at the start of Year 2. All models were run using Full Information Maximum Likelihood estimation (FIML) in Mplus version 7.11. This type of estimation accounts for missing data by using all available data for each case in estimating parameters to adjust for potential bias in the estimates resulting from missing data, and has been identified as one of the optimal ways to handle missing data in education research (Peugh & Enders, 2004).

Child (i.e., race, gender, parent years of education) and teacher (i.e., age, years of education, years of experience teaching pre-k, race) demographic variables that are commonly used as covariates in developmental and educational research (e.g., Clements et al., 2011; Gropen et al., 2017; Moilanen, Shaw, Dishion, Gardner, & Wilson, 2010; Pianta et al., 2014) were included in our models to improve the precision of the estimate of the treatment effect

(Deaton & Cartwright, 2018). Child race and child gender were aggregated to the classroom level by calculating the proportions of students in each of the relevant groups. Proportion Male, Proportion Black, and Proportion White were included in all of the predictive models. Proportion Female and the proportions of other race groups were not included in the predictive models because their collinearity with the other proportions would have prevented the models from converging. Additionally, we controlled for pre-k program type given research suggesting variability across types (Gormley, Phillips, & Gayer, 2008). We considered four different program types: Head Start, state-funded pre-k, private not for profit pre-k, and private for profit pre-k. This four-level variable was dummy coded (using private for profit as the reference group) and the three resulting dummy codes where included in analyses as covariates. Finally, we also controlled for data collection site to account for any differences across study sites.

Table 8 presents the coefficients for treatment effects on the classroom outcomes. The effect size *g* presented in the table represents the difference between the treatment and control groups divided by the standard deviation of the control group in the fall. The results show that teachers in the treatment group, as compared with teachers in the control group, had higher levels of mathematics quality, science quality, CLASS Instructional Support, and number of videos submitted across both years. The treatment group also had higher CLASS Emotional Support and Classroom Organization in Year 1, and higher proportion of time spent in mathematics and science in Year 2.

Research Question 2: Is exposure to implementation of MTP-M/S associated with improvements in children's mathematics and science outcomes during preschool relative to the comparison group?

As above, models were run at the classroom level with child-level variables aggregated to the classroom level. Child and teacher demographic characteristics were included as covariates, as were fall assessment scores, program type, and data collection site. Coefficients for treatment effects on residualized changes in the child outcomes are presented in Table 9. Note that the PAS did not have fall scores, so positive coefficients indicate that the mean of the outcome was greater in the treatment condition than in the control condition (rather than representing growth over the year). The results do not show treatment effects on child outcomes in Year 1, but do indicate greater residualized change in the treatment condition for teacher-rated mathematics and science in Year 2. Additionally, the mean of the PAS for the treatment group was higher than the mean of the PAS for the control group in Year 2.

Research Question 3: Are the associations between MTP-M/S and children's mathematics and science outcomes during preschool mediated through improved teaching practice?

Finally, mediation analyses were performed at the classroom level to determine whether effects of treatment on child outcomes (including those that showed non-significant direct associations) could be explained by the effects of treatment on measures of teacher practice. Mediation analyses were based on normal-theory tests (i.e., Sobel tests) and were conducted in Mplus using the "Model Indirect" command. We set our p-value to .01 because of the large number of mediation models being examined. The ability of each measure of teacher practice to mediate the effects on each child outcome was tested in a separate model. None of the tests were significant (ps for all indirect effects > .01), suggesting that improvements in teacher-child interactions, quality of mathematics and science teaching practice, and dosage of instruction did *not* explain why children in the treatment classrooms performed better than children in the control conditions on teacher ratings of mathematics and science and on the PAS.

Sensitivity Analyses: To what extent might Year 2 attrition be affecting our results?

We conducted additional sensitivity analyses to examine the robustness of our Year 2 results. As our first sensitivity analysis, we calculated propensity scores representing the probability of a classroom being assigned to the treatment condition based on children's skills and child, family, and teacher demographics. We then added these propensity scores to our Year 2 models (both main effect models and mediation analyses) as covariates to adjust our results for these differences (for more information on this method, see Austin, 2008, 2011). The approach the we used allowed the propensity scores to have missing values which were then addressed by FIML. None of the models involving the number of tapes converged, but the significance levels for all of our other results did not change after the addition of the propensity scores (see Appendix A for results from the models that include the propensity scores). This provides some evidence that differential attrition was not responsible for our observed effects.

As a second sensitivity analysis, we reanalyzed our Year 1 data only including those who *also* had data in Year 2. To the extent that our results are robust to attrition, we would expect these new results to parallel the original Year 1 results. For child outcomes, we replicated all of the full sample results. For classroom outcomes, we replicated the significant effects of treatment on mathematics quality, science quality, Classroom Organization, Instructional Support, and number of videos submitted, as well as the nonsignificant effect on the proportion of time engaged in science and mathematics. The effect of treatment on Emotional Support became nonsignificant (p = .12). In general, these analyses support the robustness of our Year 2 effects.

Discussion

The MTP-M/S mathematics and science curricula and implementation support system were developed to improve children's mathematics and science skills by improving the quantity

and quality of instruction in these areas. Results suggest that the MTP-M/S curricula and teacher support system have moderate to large positive effects on teacher practice. Our results also provide some evidence that MTP-M/S is associated with children's knowledge and skills in mathematics and science, but only within the context of the quasi-experimental design in teachers' second year of participation in the intervention. We did not find evidence that child outcomes were mediated by changes in the quality and quantity of mathematics and science instruction. Below, we further describe our results and their implications, limitations, and future directions.

MTP-M/S Improves Teachers' Practice

Our results suggest that, in Years 1 and 2 of teachers' participation, the intervention had large positive associations with the quality of teachers' mathematics and science instruction. In Year 1, there was also a large positive impact on the quality of teachers' Instructional Support. Although the impact was still significant in Year 2, the strength of the association between treatment condition and Instructional Support diminished (g = 1.44 and 0.46, respectively). Across both years, teachers in the MTP-M/S condition also implemented a significantly greater number of mathematics and science activities than teachers in the BaU condition. In Year 2, teachers in the MTP-M/S condition engaged in more mathematics and science instruction during live observations than teachers in the BaU condition. These results replicated some of our findings from the earlier field trial of MTP-M/S (Whittaker et al., 2016) and suggest that providing teachers with iteratively developed, research-based mathematics and science curricula, along with PD supports, can positively impact the quality and quantity of their mathematics and science instruction.

Based on research suggesting that the positive effects of early childhood mathematics curricula on teachers' mathematics instruction are not evident until teachers' second year of implementation (e.g., Clements et al., 2011; PCER, 2008), we hypothesized that we might not see positive effects of the intervention on the quality of mathematics and science instruction until Year 2 of the intervention. Counter to our hypothesis, we found significant effects on the quality of mathematics and science instruction in both Year 1 and Year 2. This could be due to the fact that we designed some of the PD supports to be embedded throughout the written curricula, so that teachers would be supported with each activity implementation from the start of the intervention.

MTP-M/S is Positively Associated with a Limited Number of Children's Mathematics and Science Knowledge and Skills

We found some evidence that, within the context of the quasi-experimental design in teachers' second year of participation in the intervention, MTP-M/S was associated with children's mathematics and science knowledge and skills. Given research suggesting that positive intervention effects on child outcomes might not be realized until teachers' second year of implementation of a new curriculum (Clements et al., 2011; PCER, 2008), it was not surprising that we did not see effects until Year 2 of the intervention. However, these findings were counter to our field trial with a smaller group of teachers and children, which showed effects after only one year of participation (Kinzie et al., 2014).

In Year 2, MTP-M/S teachers reported greater gains in children's mathematics and science skills across the pre-k year than BaU teachers. Children in MTP-M/S teachers' classrooms also scored higher on a measure of children's ability to predict, observe, reflect on, and explain scientific phenomena. These results are consistent with randomized controlled trials

suggesting that young children's learning can benefit from mathematics and science interventions that include research-based curricula and teacher PD (e.g., Clements & Sarama, 2007a; Gropen et al., 2017; Kinzie et al., 2014; Presser et al., 2015). (Note that in the current study, Year 2 is considered quasi-experimental rather than a randomized controlled trial; this is discussed further under Limitations.)

But even in Year 2, we only found small effects on child outcomes. The small effects are consistent with what have been found in other studies on the association between classroom quality and child outcomes (Burchinal, Vandergrift, Pianta, & Mashburn, 2010). However, we expected that the large impacts on mathematics and science instruction would have led to larger effects on a greater number of child outcomes. This study joins a growing group of studies that has found positive effects of curricula and PD interventions on the quality of teachers' interactions and instruction, but few significant effects on child outcomes (e.g., Morris, Mattera, & Maeir, 2016; Yoshikawa et al., 2015).

The pattern of significant associations in the current study was different than what we found in the smaller field trial of MTP-M/S (Kinzie et al., 2014). In the earlier study of MTP-M/S, we saw significant effects on direct assessments of children's mathematics outcomes. It could be that fidelity of implementation was higher or that there was less variability in implementation in the field trial. Additionally, teachers in the field trial consisted of only teachers in publicly-funded pre-k classrooms, whereas, for this study, there was a much wider range of program types. This may have led to a wider range in teachers' experience implementing mathematics and science curricula, as well as differences in program-level support for implementation. The differences in findings between the smaller field trial and this larger

efficacy trial highlight larger trends in the field around diminishing effects as interventions are implemented at greater scale (e.g., Piasta et al., 2017).

There are at least three other potential reasons for the lack of significant associations between the intervention on two of our direct assessments of children's mathematics skills (MTP-STEAM, NSPV) and one of our direct assessments of children's science skills (LENS) in this study. First, although we found higher quality interactions and instruction in treatment classrooms, the differences may not have been large enough to affect academic outcomes. For example, the mean scores for Instructional Support for treatment teachers were 2.72 and 2.51 for Year 1 and Year 2, respectively, on a seven-point scale. Supporting this very point, prior threshold analyses have suggested that Instructional Support scores of 3 or higher are associated with greater gains in children's academic skills (Burchinal et al., 2010). Similar research examining thresholds have not been conducted with the measures of mathematics and science quality, but it could be that practices in these areas were not sufficiently high to support differential gains in child outcomes.

Additionally, assignment to a treatment condition does not ensure that the intervention is being delivered as intended. In fact, research suggests that pre-k teachers may struggle with implementing curricula at high degrees of fidelity (Pianta et al., 2005). Assessing fidelity of implementation of an intervention is of critical importance in determining whether the null effects of an intervention are due to a failure of implementation, or failure of the intervention (e.g., Abry, Rimm-Kaufman, Larsen, & Brewer, 2013; Hulleman & Cordray, 2009). Thus, before conclusions are drawn about the efficacy of the intervention, replication is need, as are assessments of fidelity, which can be examined by evaluating the components of adherence,

dosage, quality of delivery, participant responsiveness, and program differentiation (Berkel, Mauricio, Schoenfelder, & Sandler, 2011; Durlak & Dupre, 2008).

There was also a great deal of variability in teachers' usage of PD supports. Teachers' usage of online supports ranged from one minute to over 36 hours. There was also variability in the number of workshops that teachers attended. It could be that treatment teachers who engaged more with the PD supports were able to implement the curricula with higher quality and fidelity, which ultimately led to greater gains in children's mathematics and science outcomes. These are follow-up questions about the treatment group that we will explore in future analyses. We will also examine whether there are certain teacher characteristics, attitudes, and beliefs that might lead teachers to engage more or less with PD supports, in an effort to better understand how to support teachers' engagement with them.

Finally, we did not follow children into the kindergarten year. It has been suggested that some interventions produce "sleeper effects", where the impact does not appear until after the conclusion of the intervention (Morris et al., 2016; Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010). It is possible that increased time and high-quality instruction in the pre-k year laid the foundation for greater mathematics and science skill development that would become apparent once children entered kindergarten, especially if the kindergarten classroom builds on children's existing knowledge base (Engel, Claessens, & Finch, 2013).

Changes in Individual Teacher Practices Were Not the Mechanism through which the Intervention had Effects on Child Outcomes.

In our theory of change, we hypothesized that the intervention would directly impact both teachers' practice and children's outcomes, and that there would be indirect effects of the intervention on students' mathematics and science knowledge and skills through positive effects

on the quality of teachers' interactions and instruction. Although the intervention was associated with increases in the quality of teachers' interactions and instruction, and a limited number of child outcomes, our mediation analysis suggested that these isolated components of teachers' practice were not the mechanisms through which the intervention was associated with student outcomes. It is possible that our approach to examining mediation was too global. In their evaluation of the Building Blocks curriculum, Clements and colleagues (2011) examined the mediational role of instructional practices on child outcomes. They found that very specific components of the COEMET (e.g., availability of computers in the classroom) partially mediated the effects of the curriculum on outcomes, but they did not find the global mathematics quality score to be a significant mediator. Differences in study design and in our use of the COEMET precluded us from testing the same mediators, but the findings do raise questions about whether certain specific practices may transmit intervention effects more than global quality of instruction.

Another potential explanation is that it may not be these factors in isolation, but *combinations* of quality and dosage that lead to improved outcomes. Very low, largely non-significant correlations between dosage and quality suggest that our sample included wide variability in implementation: Some teachers may have taught many lessons but did so with poor quality, and others taught few lessons but of high quality. A more complex examination of implementation was beyond the scope of this paper but is of interest in future work. It may be possible to identify a sub-group of teachers who were "high" and "low" implementers and determine how these combinations of dosage and quality were associated with gains.

Finally, it could be that the intervention impacted some other unmeasured teacher practices that resulted in changes in child outcomes. Research suggests that teacher language

input, specifically the amount and type of “math talk” is a key mechanism for fostering children’s mathematics learning (e.g., Klibanoff et al., 2006). This is one example of an unmeasured factor that may have been impacted by the intervention and led to changes in child outcomes, although there are undoubtedly others. With fairly consistent findings of weak to null associations between measures of early childhood quality and child outcomes (e.g., Burchinal, 2017), it will be important for the field of early childhood to continue to examine “high-leverage” teacher practices that lead to improvements in child outcomes.

Limitations

There are several limitations to this study that warrant attention. There was substantial attrition from Year 1 to Year 2 of the study. In Year 1, we examined and found no significant differences in sample charateristics and baseline child outcome variables between conditions at either timepoint. However, in Year 2, we were not able to establish baseline equivalence between conditions for two of our measures. In Year 2, children in control classrooms started the study with higher scores on the MTP-STEAM (d=-0.32) and NSPV (d=-0.34) than children in intervention classrooms. It could also be that there were other unmeasured factors that differed, causing threats to the validity of study findings. For example, we did not have a measure of teacher practice at baseline, and therefore could not test for equivalence among the intervention and control conditions in teacher practices. This is an especially important consideration given that we only found significant associations between the intervention and child outcomes in Year 2. Additionally, for the child outcomes where we did find significant effects, two were teacher report measures. Although teachers can be valuable reporters of children’s skills given the amount of time spent with them in the classroom setting, some research suggests that they may lack validity due in part to bias toward or against individual children and because significant

variability in teachers' report of children's skills can be explained by teacher and student characteristics (e.g., Furnari, Whittaker, Kinzie, & DeCoster, 2016; Waterman, McDermott, Fantuzzo, & Gadsden, 2012).

The direct assessment on which we found significant treatment effects, the PAS (Gropen et al., 2017), was administered at post-test only, so we were not able to control for baseline scores and possible differences across treatment groups at pre-test. In addition, the PAS had lower-than-desirable internal consistency, which may have reduced our ability to detect effects. There also was some level of alignment between the MTP-M/S curriculum and the content included in the PAS, which may have positively biased students' performance in the treatment condition. The MTP-M/S curriculum includes two activities on floating and sinking (but does not include activities on water level). However, as other researchers have noted (Gropen et al., 2017), this is a common activity across preschool classrooms, so it could be that children in control classrooms also had similar experience with this topic.

Third, compared to the area of literacy, there are relatively few existing measures of the quality of teachers' instruction in mathematics and science (Brenneman et al., 2011; Halle, Whittaker, & Anderson, 2010) and of children's knowledge and skills in these areas (NRC, 2008). In this study, we adapted several existing measures including the Short TEAM (Weiland et al., 2012), and the Academic Rating Scales from the ECLS-K (National Center for Educational Statistics, 1998, 2011). We also developed our own observational measure of science quality (Vitiello et al., 2018), as there was not an existing published and validated measure available. Although these measures correlated in ways that we would expect with other validated measures, additional work needs to be done to determine their psychometric properties across early childhood settings.

Finally, we used the number of tapes that teachers submitted as one of our measures of dosage of mathematics and science instruction. The field of implementation science has recently drawn attention to the importance of measuring and understanding how much of an intervention participants actually receive (Wasik, Mattera, Lloyd, & Boller, 2013). But there is currently a lack of consensus about the best approach to use in measuring dosage. Although others have also used this approach as a measure of intervention implementation (e.g., Pianta et al., 2008), it is possible that this method confounds aspects of both dosage and adherence (i.e., conducting the videotaping and returning those tapes). As the field of implementaiton science moves forward, researchers need to continue to evaluate and determine the most valid approaches to measuring intervention dosage.

Conclusion

There is growing recognition that high-quality mathematics and science instruction in the early years can significantly improve mathematics and scientific understanding and the capacity for complex and abstract thought (Eshach & Fried, 2005; NRC, 2001, 2007b). However, several challenges impede efforts to improve outcomes among young children including relatively few evidence-based curricula in these domains (e.g., NCQTL, 2015) and limited formal education and training opportunities for teachers to support their mathematics and science instruction (Ginsburg et al., 2006; Institute of Medicine & NRC, 2015). This study demonstrates that curricular interventions can increase the quality and quantity of teachers' mathematics and science instruction. There was also evidence in the quasi-experimental design of Year 2 that the intervention may have positive benefits for children's mathematics and science knowledge and skills. However, more work is needed to better understand the mechanisms of change that lead to gains in children's mathematics and science outcomes in the pre-k year. With this

understanding, efforts can be made to bolster the factors that produce those gains, and increase their effects.

References

Abry, T., Rimm-Kaufman, S. E., Larsen, R. A., & Brewer, A. J. (2013). The influence of fidelity of implementation on teacher-student interaction quality in the context of a randomized controlled trial of the Responsive Classroom approach. *Journal of School Psychology, 51*, 437–453. doi: 10.1016/j.jsp.2013.03.001

Austin, P. C. (2008). Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score. *Pharmacoepidemiology and Drug Safety, 17*, 1202–1217. http://dx.doi.org/10.1002/pds.1673

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*, 399–424. http://dx.doi.org/10.1080/ 00273171.2011.568786

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy: A tool for curriculum reform*. New York: Oxford University Press.

Backscheider, A. G., Shatz, M., & Gelman, S. A. (1993). Preschoolers' ability to distinguish living kinds as a function of regrowth. *Child Development, 64*, 1242-1257. doi: 10.1111/j.1467-8624.1993.tb04198.x

Baroody, A. J. (1992). The development of preschoolers' counting skills and principles. In J. Bideau, C. Meljac, & J. P. Fischer (Eds.), *Pathways to number* (pp. 99–126). Hillsdale, NJ: Erlbaum.

Baroody, A. J., Lai, M. L., & Mix, K. S. (2006). The development of number and operation sense in early childhood. In O. Saracho & B. Spodek (Eds.), *Handbook of research on the education of young children* (pp. 187–221). Mahwah, NJ: Erlbaum.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*, 23–33. doi:10.1007/s11121-010-0186-1

Bowman, B. T., Donovan, M. S., & Burns, M. S. (2001). *Eager to Learn: Educating our preschoolers*. Washington, DC: National Research Council.

Brenneman, K., Boller, K., Atkins-Burnett, S., Stipek, D., Forry, N., Ertle, B., … Schultz, T. (2011). Measuring the quality of early childhood math and science curricula and teaching. In Martha Zaslow, Ivelisse Martinez-Beck, Kathryn Tout, & Tamara Halle (Eds.), *Quality measurement in early childhood settings.* Baltimore, MD: Paul H. Brookes.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32-42. doi: 10.3102/0013189X018001032

Brown J., & Greenfield, D.B. (2010). *Early childhood hands-on science (ECHOS)*. Institute of Education Sciences Goal 3 RCT efficacy Grant: R305A100275.

Bryant, P. (1995). Children and arithmetic. *Journal of Child Psychology and Psychiatry, 36*, 3–32. doi: 10.1111/j.1469-7610.1995.tb01654.x

Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, *12*, 3-9. doi:10.1111/cdep.12260

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*, 166-176. doi: 10.1016/j.ecresq.2009.10.004

Carpenter, T. P. & Lehrer, R. (1999) Teaching and learning mathematics with understanding. In E. Fennema, & T. A. Romberg (Eds.), *Mathematics classrooms that promote understanding* (pp. 19 – 32). New York, NY: Routledge

Chalufour, I. & Worth, K. (2003) *The Young Scientist Series: Discovering nature with young children*. St. Paul, MN: Redleaf Press.

Chalufour, I. & Worth, K. (2004) *The Young Scientist Series: Building structures with young children*. St. Paul, MN: Redleaf Press.

Chalufour, I. & Worth, K. (2005) *The Young Scientist Series: Exploring water with young children*. St. Paul, MN: Redleaf Press.

Cicchetti, D., Bronen, R., Spencer, S., Haut, S., Berg, A., Oliver, P., & Tyrer, P. (2006). Rating scales, scales of measurement, issues of reliability: Resolving some critical issues for clinicians and researchers. *The Journal of Nervous and Mental Disease, 194*, 557-564. doi: 10.1097/01.nmd.0000230392.83607.c5

Clements, D. H. (2004a). Geometric and spatial thinking in early childhood education. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in Mathematics: Standards for early childhood mathematics education* (pp. 267-297). Mahwah, NJ: Lawrence Erlbaum.

Clements, D. H. (2004b). Major themes and recommendations. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 7-72). Mahwah, NJ: Lawrence Erlbaum.

Clements, D. H., & Sarama, J. (2007a). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education, 38*, 136–163.

Clements, D. H., & Sarama, J. (2007b). *Building Blocks-SRA real math teacher's edition, Grade PreK*. Columbus, OH: SRA/McGraw-Hill.

Clements, D. H., & Sarama, J. (2007c). Early childhood mathematics learning. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 461-555). New York: Information Age Publishing.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Education Research Journal, 45*, 443-494. doi: 10.3102/0002831207312908

Clements, D. H., & Sarama, J. (Eds.). (2004). *Engaging young children in Mathematics: Standards for early childhood mathematics education*. Mahwah, NJ: Lawrence Erlbaum.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42*, 127-166.

Clements, D. H., Swaminathan, S., Hannibal, M. A. Z., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education, 30*, 192–212. doi: 10.2307/749610

Cognition and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19*, 2-10. doi: 10.3102/0013189X019006002

Conezio, K., & French, L. (2002). Science in the preschool classroom. *Young Children, 57*(5), 12-18.

Davis, E. A., & Krajcik, J. S. (2005). Desgining educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3-14. doi: 10.3102/0013189X034003003

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine, 210*, 2-21.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350. doi:10.1007/s10464-008-9165-0

Engel, M., Claessens, A., & Finch, M. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content in mathematics and student knowledge in kindergarten. *Educational Evaluation and Policy Administration, 35,* 157–178. doi: 10.3102/0162373712461850

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology, 14,* 315–336. doi: 10.1007/s10956-005-7198-9

Forbes, C. T., & Davis, E. A. (2010). Curriculum design for inquiry: Preservice elementary teachers' mobilization and adaptation of science curriculum materials. *Journal of Research in Science Teaching, 47*, 820-839. doi: 10.1002/tea.20379

French, L. A. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Childhood Research Quarterly, 19*, 138-149. doi: 10.1016/j.ecresq.2004.01.004

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development, 60*, 1158-1171. doi: 10.2307/1130790

Furnari, E. C., Whittaker, J., Kinzie, M., & DeCoster, J. (2016). Factors associated with accuracy in prekindergarten teacher ratings of students' mathematics skills. *Journal of Psychoeducational Assessment*, *35*, 410-423.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945. doi: 10.3102/00028312038004915

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, *19*, 150-158. doi: 10.1016/j.ecresq.2004.01.009

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard U. Press.

Ginsburg, H. P., & Ertle, B. (2008). Knowing the mathematics in early childhood mathematics. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 45-65). Charlottes, NC: Information Age Publishing.

Ginsburg, H. P., & Golbeck, S. L. (2004). Thoughts on the future of research on mathematics and science learning and education. *Early Childhood Research Quarterly*, *19*, 190–200. doi: 10.1016/j.ecresq.2004.01.013

Ginsburg, H. P., Greenes, C., & Balfanz, R. (2003). *Big Math for Little Kids: Classroom Set*: Dale Seymour.

Ginsburg, H. P., Kaplan, R. G., Cannon, J., Cordero, M. I., Eisenband, J. G., Galanter, M., & Morgenlander, M. (2006). Helping early childhood educators to teach mathematics. In M. Zaslow & I. Martinez-Beck (Eds.), *Critical issues in early childhood professional development* (pp. 171–202). Baltimore, MD: Paul H. Brookes.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report, 22*(1).

Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science, 320*, 1723–1724. doi: 10.1126/science.1156019

Greenes, C., Ginsburg, H. P., & Balfanz, R. (2004). Big Math for Little Kids. *Early Childhood Research Quarterly, 19*, 159-166. doi: 10.1016/j.ecresq.2004.01.010

Greenfield, D. B., Dominguez, X., Fuccillo, J. M., Maier, M. F., Greenberg, A. C., & Penfield, R. (2009). *Development of an IRT-based direct assessment of preschool science*. Paper presented at the Biennial meeting of the Society for Research in Child Development, Denver, CO.

Greenfield, D. B., Dominguez, X., Greenberg, A., Fuccillo, J., & Maier, M. (2011). *Assessing science readiness in young low-income preschool children*. Paper presented at the Biennial meeting of the Society for Research in Child Development, Montreal.

Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the preschool classroom: A programmatic research agenda to improve science. *Early Education and Development, 20*, 238-264. doi: 10.1080/10409280802595441

Gropen, J., Kook, J. F., Hoisington, C., & Clark-Chiarelli, N. (2017). Foundations of science literacy: Efficacy of a preschool professional development program in science on classroom instruction, teachers' pedagogical content knowledge, and children's observations and predictions. *Early Education and Development, 28*, 607-631. doi: 10.1080/10409289.2017.1279527.

Gutiérrez, R. (2008). A "gap-gazing" fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education, 39*, 357-364.

Halle, T., Whittaker, J. V., & Anderson, R. (2010). *Quality in early childhood care and education settings: A compendium of measures, second edition*. Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Harlan, J. D., & Rivkin, M. S. (2004). *Science experiences for the early childhood years: An integrated affective approach* (8th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hofer, K. G., Lipsey, M. W., Dong, N., & Farran, D. C. (2013). *Results of the Early Math Project – scale-up cross-site results* (Working Paper). Nashville, TN: Vanderbilt University, Peabody Research Institute.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88–110. doi: 10.1080/19345740802539325

Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child Development, 67*, 2823-2840. doi: 10.1111/j.1467-8624.1996.tb01890.x

Institute of Medicine & National Research Council. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. Washington, DC: The National Academies Press. doi: 10.17226/19401

Justice, L. M., & Pullen, P. C. (2003). Promising interventions for promoting emergent literacy skills. *Topics in Early Childhood Special Education, 23*, 99–113. doi: 10.1177/02711214030230030101

Kemple, J., Corrin, W., Nelson, E., Salinger, T., Herrmann, S., and Drummond, K. (2008). *The Enhanced Reading Opportunities study: Early impact and implementation findings* (NCEE 2008-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Kinzie, M. B., Whittaker, J. V., McGuire, P., Lee, Y., & Kilday, C. R. (2015). Research on curricular development for pre-kindergarten mathematics & science. *Teachers College Record, 117*(7), 1-40.

Kinzie, M. B., Whittaker, J. V., Williford, A. P., DeCoster, J., McGuire, P., Lee, Y., & Kilday, C. R. (2014). MyTeachingPartner-Math/Science pre-kindergarten curricula and teacher supports: Associations with children's math and science learning. *Early Childhood Research Quarterly*, *29,* 586-599. doi: 10.1016/ j.ecresq.2014.06.007

Klahr, D., Zimmerman, C., & Jirout, J., (2011) Educational interventions to advance children's scientific thinking. *Science*, *333,* 971-975. doi: 10.1126/science.1204528

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1,* 155–178. http://doi.org/10.1080/19345740802114533

Klein, A., Starkey, P., & Ramirez, A. (2002). *Pre-K Mathematics Curriculum*. Glendale, IL: Scott Foresman.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk". *Developmental Psychology, 42,* 59-69. doi: 10.1037/0012-1649.42.1.59

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674–689. doi: 10.1037/0033-295X.96.4.674

Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology, 53*, 72-103. doi: 10.1016/S0022-0965(05)80005-0

Lai, E. R. (2011). *Motivation: A literature review*. Retrieved from: https://images.pearsonassessments.com/images/tmrs/Motivation_Review_final.pdf

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., . . . Howes, C. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language and social skills. *Child Development, 79*, 732-749. doi: 10.1111/j.1467-8624.2008.01154.x

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology, 67*, 267-277.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, *47*, 474–496. doi: 10.1002/tea.20347

Moilanen, K. L., Shaw, D. S., Dishion, T. J., Gardner, F., & Wilson, M. (2010). Predictors of longitudinal growth in inhibitory control in early childhood. *Social Development*, *19*, 326–347. doi: 10.1111/j.1467-9507.2009.00536.x

Moreno, R., & Ortegano-Layne, L. (2008). Do classroom exemplars promote the application of principles in teacher education? A comparison of videos, animations, and narratives. *Educational Technology Research & Development, 56*, 449-465. doi: 10.1007/s11423-006-9027-0

Morris, P. A., Mattera, S. K., & Maier, M. F. (2016). *Making pre-k count: Improving math instruction in New York City*. Retrieved from MDRC website: http://www.mdrc.org/sites/default/files/Making_Pre-K_Count_FR.pdf

National Association for the Education of Young Children & National Council of Teachers of Mathematics. (2002). *Early childhood mathematics: Promoting good beginnings.* Retrieved from National Association for the Education of Young Children website: https://www.naeyc.org/files/naeyc/file/positions/psmath.pdf

National Center for Education Statistics. (1998). *Academic Rating Scale*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from: https://nces.ed.gov/ecls/pdf/kindergarten/springteachersABC.pdf

National Center for Education Statistics. (2011). *Academic Rating Scale*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from: https://nces.ed.gov/ecls/pdf/kindergarten2011/Spring_K_Classroom_Teacher_Child_Level.pdf

National Center on Quality Teaching and Learning. (2015). *Mathematics preschool curriculum consumer report*. Retrieved from the Head Start, Administration for Children and Families, U.S. Department of Health and Human Services website: https://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/docs/math-preschool-curriculum-report.pdf

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA: The National Council of Teachers of Mathematics, Inc.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Retrieved from the U.S. Department of Education website: http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

National Research Council. (1996). *National science education standards*. Washington, DC: The National Academies Press. doi: 10.17226/4962

National Research Council (2000a) *Eager to learn: Educating our preschoolers*. Washington, DC: The National Academies Press. doi: 10.17226/9745

National Research Council (2000b) *How people learn: Brain, mind, experience, and school: Expanded Edition*. Washington, DC: The National Academies Press. doi: 10.17226/9853.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: The National Academies Press. doi: 10.17226/9822

National Research Council. (2005). *Mathematical and scientific development in early childhood: A workshop summary*. Washington, DC: The National Academies Press. doi: 10.17226/11178

National Research Council. (2006). *National science education standards.* Washington, DC: National Academies Press.

National Research Council. (2007a). *Enhancing professional development for teachers: Potential uses of information technology: Report of a workshop*. Washington, DC: The National Academies Press. doi: 10.17226/11995

National Research Council. (2007b). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: The National Academies Press: doi:10.17226/11625.

National Research Council. (2008). *Early childhood assessment: Why, what, and how.* Washington, DC: The National Academies Press. doi: 10.17226/12446

National Research Council. (2009). *Mathematics learning in early childhood*: *Paths toward excellence and equity*. Washington, DC: The National Academies Press. doi: 10.17226/12519

Nayfeld, I., Brenneman, K., & Gelman, R. (2011). Science in the classroom: Finding a balance between autonomous exploration and teacher-led instruction in preschool settings. *Early Education and Development, 22*, 970-988. doi: 10.1080/10409289.2010.507496

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*, 525-556.

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., . . . Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. A consensus statement*. Washington, DC: The Brookings Institution.

Piaget, J. & Inhelder, B. (1969) *The psychology of the child.* New York: Basic Books.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*, 144-159. doi: 10.1207/s1532480xads0903_2

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS), Pre-K*. Baltimore, MD: Paul H. Brookes.

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*, 431-451. doi: 10.1016/j.ecresq.2008.02.001

Pianta, R. C., DeCoster, J., Cabell, S., Burchinal, M., Hamre, B. K., Downer, J., … Howes, C. (2014). Dose-response relations between preschool teachers' exposure to components of professional development and increases in quality of their interactions with children. *Early Childhood Research Quarterly*, *29*, 499–508. doi: 10.1016/j.ecresq.2014.06.001

Piasta, S. B., Justice, L. M., O'Connell, A. A., Mauck, S. A., Weber-Mayrer, M., Schachter, R. E., . . . Spear, C. F. (2017). Effectiveness of large-scale, state-sponsored language and literacy professional development on early childhood educator outcomes. *Journal of Research on Educational Effectiveness, 10*(2), 354–378.

Piasta, S. B., Logan, J. A. R., Pelatti, C. Y., Capps, J. L., & Petrill, S. A. (2015). Professional development for early childhood educators: Efforts to improve math and science learning opportunities in early childhood classrooms. *Journal of Educational Psychology, 107*, 407-422. doi: http://dx.doi.org/10.1037/a0037621

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness* (NCER 2008-2009). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

Presser, A. L., Clements, M., Ginsburg, H., & Ertle, B. (2015). Big Math for Little Kids: The effectiveness of a preschool and kindergarten mathematics curriculum. *Early Education and Development, 26*, 399-426. doi: 10.1080/10409289.2015.994451

Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2001). *Emergent Academic Snapshot*. Unpublished instrument. University of California, Los Angeles.

Ruffman, T., Olson, D. R., Ash, T., & Keen, T. (1993). The ABCs of deception: Do young children understand deception in the same way as adults? *Developmental Psychology, 29,* 74-87. doi: 10.1037/0012-1649.29.1.74

Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development, 64*, 1617-1636. doi: 10.1111/j.1467-8624.1993.tb04203.x

Sarama, J., & Clements, D. H. (2003). Building blocks of early childhood mathematics. *Teaching Children Mathematics, 9*, 480-484.

Sarama, J., & Clements, D. H. (2007). *Manual for Classroom Observation of Early Mathematics: Environment and Teaching, version 3*. Unpublished instrument. Buffalo, NY: Research and Assessment Center.

Sarama, J., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2016). Professional development in early mathematics: Effects of an intervention based on learning trajectories on teachers' practices. *Nordic Studies in Mathematics Education, 21*(4), 29–55.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*, 20-26.

Snow, K. (2011). *Research news you can use: Debunking the play vs. learning dichotomy*. Retrieved from http://www.naeyc.org/content/research-news-you-can -use-play-vs-learning

Sousa, D. A. (2008). *How the brain learns mathematics*. Thousand Oaks, CA: Corwin Press.

Sodian, B., Zaitchik, D., & Carey, S. (1991) Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62*, 753-766. doi: 10.1111/j.1467-8624.1991.tb01567.x

Springer, K., & Keil, F. C. (1991). Early differentiation of causal mechanisms appropriate to biological and nonbiological kinds. *Child Development, 62*, 767-781. doi: 10.1111/j.1467-8624.1991.tb01568.x

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly, 19*, 99-120. doi: 10.1016/j.ecresq.2004.01.002

Stock, P., Desoete, A., & Roeyers, H. (2009). Mastery of the counting principles in toddlers: A crucial step in the development of budding arithmetic abilities? *Learning and Individual Differences, 19*, 419-422. doi: 10.1016/j.lindif.2009.03.002

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. *Child Development*, *81*, 737–756. doi: 10.1111/j.1467-8624.2010.01431.x

Vitiello, V. E., Whittaker, J. V., Kinzie, M. B., Mulcahy, C., & Helferstay, L. (2018, online first). Reliability and validity of the Pre-K Science Observation Measure (P-SOM). *Early Education and Development*. doi: 10.1080/10409289.2018.1544814

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Pyschology, 98*, 63-74. doi: 10.1037/0022-0663.98.1.63

Wasik, B. A., Mattera, S. K., Lloyd, C. M., & Boller, K. (2013). *Intervention dosage in early childhood care and education: It's complicated* (OPRE Research Brief OPRE 2013-15). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—Or whose score is it anyway? *Early Childhood Research Quarterly*, *27*, 46-54.

Weiland, C., Wolfe, C. B., Hurwitz, M. D., Clements, D. H., Sarama, J. H., & Yoshikawa, H. (2012). Early mathematics assessment: Validation of the short form of a prekindergarten and kindergarten mathematics measure. *Educational Psychology, 32*, 311-333. doi:10.1080/01443410.2011.654190

What Works Clearinghouse (2017). Standards Handbook Version 4.0. Downloaded from https://files.eric.ed.gov/fulltext/ED577036.pdf on November 29, 2018.

What Works Clearinghouse (2018). Assessing attrition bias. Downloaded from https://ies.ed.gov/ncee/wwc/Document/243 on November 29, 2018.

Whittaker, J. V., Kinzie, M. B., Williford, A., & DeCoster, J. (2016). Effects of MyTeachingPartner-Math/Science on teacher-child interactions in prekindergarten classrooms. *Early Education and Development*, *27*, 110-127. doi: 10.1080/10409289.2015.1047711

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155-193. doi: 10.1016/0010-0277(90)90003-3

Yoon, J., & Onchwari, J. A. (2006). Teaching young children science: Three key points. *Early Childhood Education Journal*, *33*, 419-423. doi: 10.1007/s10643-006-0064-4

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M. C., Weiland, C., … Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology*, *51*, 309–322. doi: 10.1037/a0038785

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). Emerging research on early childhood professional development. In S. B. Neuman, & M. L. Kamil (Eds.), *Preparing Teachers for the Early Childhood Classroom* (19 – 48). Baltimore, MD: Paul H. Brookes.

Zucker, A. A., Tinker, R., Staudt, C., Mansfield, A., & Metcalf, S. (2008). Learning science in grades 3-8 using probeware and computers: Findings from the TEEMSS II project. *Journal of Science Education and Technology*, *17*, 42–48. https://doi.org/10.1007/s10956-007-9086-y

Zur, O., & Gelman, R. (2004). Young children can add and subtract by predicting and checking. *Early Childhood Research Quarterly*, *19*, 121-137. doi: 10.1016/j.ecresq.2004.01.003

Table 1

Descriptive Characteristics of Study Variables

	Children *N*=1371
Child/Family Demographics	*M*(*SD*) or *n* (%)
Race	
Black/African American	400 (29%)
White	733 (53%)
Other	238 (17%)
Gender	
Male	686 (50%)
Child age in months	53.78 (3.84)
Parent years of education	14.45 (2.48)
	Teachers *N*=140
Teacher Demographics	*M(SD)* or *n* (%)
Race	
Black/African American	26 (19%)
White	96 (69%)
Other	18 (12%)
Gender	
Female	129 (92%)
Male	3 (2%)
Unreported	8 (6%)
Teacher age in years	37.54 (10.87)
Years of education	15.35 (1.79)
Years of experience teaching pre-k	6.92 (7.09)

Table 2

Teacher Practice and Child Outcomes, by Condition and Year

	Control (BaU)		Treatment (MTP-M/S)	
Teacher Practice	Year 1 *Mean (SD)* *n* = 140	Year 2 *Mean (SD)* *n* = 75	Year 1 *Mean (SD)* *n* = 140	Year2 *Mean (SD)* *n* = 75
Math quality	1.52 (0.26)	1.64 (0.37)	2.15 (0.33)	2.22 (0.39)
Science quality	3.71 (0.74)	3.46 (0.61)	4.85 (0.45)	4.66 (0.51)
CLASS Emotional Support	4.87 (0.44)	4.88 (0.50)	5.05 (0.37)	4.86 (0.35)
CLASS Classroom Organization	5.18 (0.49)	5.32 (0.39)	5.42 (0.39)	5.37 (0.36)
CLASS Instructional Support	2.08 (0.45)	2.3 (0.42)	2.72 (0.40)	2.51 (0.36)
Proportion of time engaged in math	0.10 (0.09)	0.06 (0.06)	0.09 (0.07)	0.1 (0.09)
Proportion of time engaged in science	0.04 (0.05)	0.04 (0.05)	0.06 (0.07)	0.09 (0.10)
Number of videos submitted	43.90 (30.34)	43.63 (32.32)	72.96 (43.35)	70.31 (42.17)
Child Outcomes				
ARS Math				
Pre	3.07 (0.82)	2.85 (0.84)	3.08 (0.87)	2.83 (0.87)
Post	4.09 (0.56)	4.13 (0.51)	4.05 (0.63)	4.36 (0.38)
NSPV				
Pre	25.59 (4.21)	25.67 (3.65)	25.35 (4.62)	24.44 (3.57)
Post	30.11 (3.30)	30.39 (2.90)	30.52 (3.66)	30.24 (3.43)
MTP-STEAM				
Pre	14.04 (3.22)	14.35 (3.02)	14.03 (3.44)	13.49 (2.45)
Post	16.89 (2.73)	17.42 (2.52)	16.73 (2.96)	16.95 (2.44)
ARS Science				
Pre	2.85 (0.93)	2.61 (0.91)	2.86 (0.86)	2.57 (0.85)
Post	3.77 (0.76)	3.98 (0.61)	3.87 (0.65)	4.17 (0.48)
LENS				
Pre	1.15 (0.65)	1.11 (0.73)	1.20 (0.69)	1.09 (0.60)
Post	1.74 (0.67)	1.79 (0.64)	1.83 (0.60)	1.84 (0.58)
PAS				
Pre	—	—	—	—
Post	.68 (.10)	.66 (.08)	.68 (.09)	.68 (.09)

Table 3

Respondents by Condition and Timepoint

	Control *n (%)*	Intervention *n (%)*	Differential Attrition (%)	Overall Attrition (%)
Full Sample	79	77		
Year 1 Fall	70 (88.6%)	70 (90.9%)	2.3%	10.3%
Year 1 Spring	62 (78.5%)	64 (83.1%)	4.6%	19.2%
Year 2 Fall	35 (44.3%)	40 (51.9%)	7.6%	51.9%
Year 2 Spring	32 (40.5%)	37 (48.1%)	7.6%	55.8%

Table 4

Baseline Equivalence Effect Sizes by Timepoint

	Baseline Equivalence Effect Size *d*			
	Fall Year 1	Spring Year 1	Fall Year 2	Spring Year 2
ARS Math	0.01	0.04	-0.02	-0.01
NSPV	-0.06	-0.08	-0.34	-0.39
MTP-STEAM	0.00	0.04	-0.32	-0.41
ARS Science	0.07	0.06	-0.02	-0.12
Lens	0.01	0.05	-0.04	-0.09

Note: Effect sizes are calculated as (Treatment mean – Control mean) / pooled standard deviation and represent the difference in Fall Year 1 scores for students still in the sample at the time point indicated by the column header.

Table 5

Bivariate Correlations among the Year 1 Teacher Outcomes

	1	2	3	4	5	6	7	8
1. Mathematics quality	—							
2. Science quality	.67*	—						
3. CLASS Emotional Support	.33*	.43*	—					
4. CLASS Classroom Organization	.35*	.54*	.77*	—				
5. CLASS Instructional Support	.65*	.67*	.54*	.57*	—			
6. Proportion of time engaged in math	.01	-.02	-.13	-.10	-.04	—		
7. Proportion of time engaged in science	.14	.17	.01	.04	.08	-.07	—	
8. Number of videos submitted	.28*	.22	.10	.16	.23	.11	.28*	—

* $p < .01$

Table 6

Bivariate Correlations among the Fall Year 1 Child Outcomes and Spring PAS

	1	2	3	4	5	6
1. ARS Math	—					
2. NSPV	.54	—				
3. MTP-STEAM	.50	.77	—			
4. ARS Science	.91	.51	.45	—		
5. Lens	.56	.75	.78	.53	—	
6. Spring PAS (not assessed in Fall)	.51	.54	.51	.49	.63	—

Note. All correlations are significant ($p < .001$).

Table 7

Bivariate Correlations of the Fall Year 1 Child Outcomes and Spring PAS with the Year 1 Classroom Outcomes

	ARS Math	NSPV	MTP-STEAM	ARS Science	LENS	Spring PAS
1. Mathematics quality	-.05	.05	.04	-.09	.00	-.06
2. Science quality	-.12	-.11	-.10	-.17	-.09	-.08
3. CLASS Emotional Support	.11	.11	.20	.13	.16	.05
4. CLASS Classroom Organization	-.06	.04	.14	-.06	.13	.08
5. CLASS Instructional Support	.06	.13	.15	.02	.13	.06
6. Proportion of time engaged in math	-.08	.08	-.01	-.07	.04	.06
7. Proportion of time engaged in science	.10	.17	.20	.10	.15	.13
8. Number of videos submitted	.12	.09	.17	.07	.17	.18

Note: All correlations are not significant ($p > .01$).

Table 8

Standardized Coefficients Relating Treatment to Classroom Outcomes Controlling for Covariates

Year 1 Outcome	*beta*	*SE{beta}*	*p*-value	effect size *g*
Mathematics quality	.723	.043	< .001	2.43
Science quality	.685	.049	< .001	1.51
CLASS Emotional Support	.220	.082	.007	0.41
CLASS Classroom Organization	.269	.081	.001	0.5
CLASS Instructional Support	.610	.055	< .001	1.44
Proportion of time engaged in math	-.098	.087	.26	-0.17
Proportion of time engaged in science	.163	.084	.05	0.4
Number of videos submitted	.357	.071	< .001	0.94
Year 2 Outcome				
Mathematics quality	.649	.087	< .001	1.68
Science quality	.765	.076	< .001	2.04
CLASS Emotional Support	-.066	.122	.59	-0.11
CLASS Classroom Organization	.063	.131	.63	0.12
CLASS Instructional Support	.242	.119	.04	0.46
Proportion of time engaged in math	.247	.112	.03	0.62
Proportion of time engaged in science	.255	.112	.02	0.84
Number of videos submitted	.395	.096	< .001	0.98

Note. Treatment was coded 1 = Treatment, 0 = Control, so positive coefficients represent more positive changes in the treatment condition than in the control condition.

Table 9

Standardized Coefficients Relating Treatment to Spring Child Outcomes Controlling for Fall Child Outcomes and Covariates

Year 1 Outcome	*beta*	*SE{beta}*	*p*-value	effect size *g*
ARS Math	-.063	.072	.39	-0.09
NSPV	.085	.066	.20	0.14
MTP-STEAM	-.034	.061	.57	-0.06
ARS Science	.019	.069	.78	0.03
LENS	.053	.051	.30	0.10
PAS	-.030	.069	.66	-0.06
Year 2 Outcome				
ARS Math	.238	.087	.006	0.26
NSPV	.112	.088	.21	0.17
MTP-STEAM	.041	.072	.57	0.06
ARS Science	.212	.087	.01	0.25
LENS	.100	.069	.15	0.2
PAS	.220	.088	.01	0.39

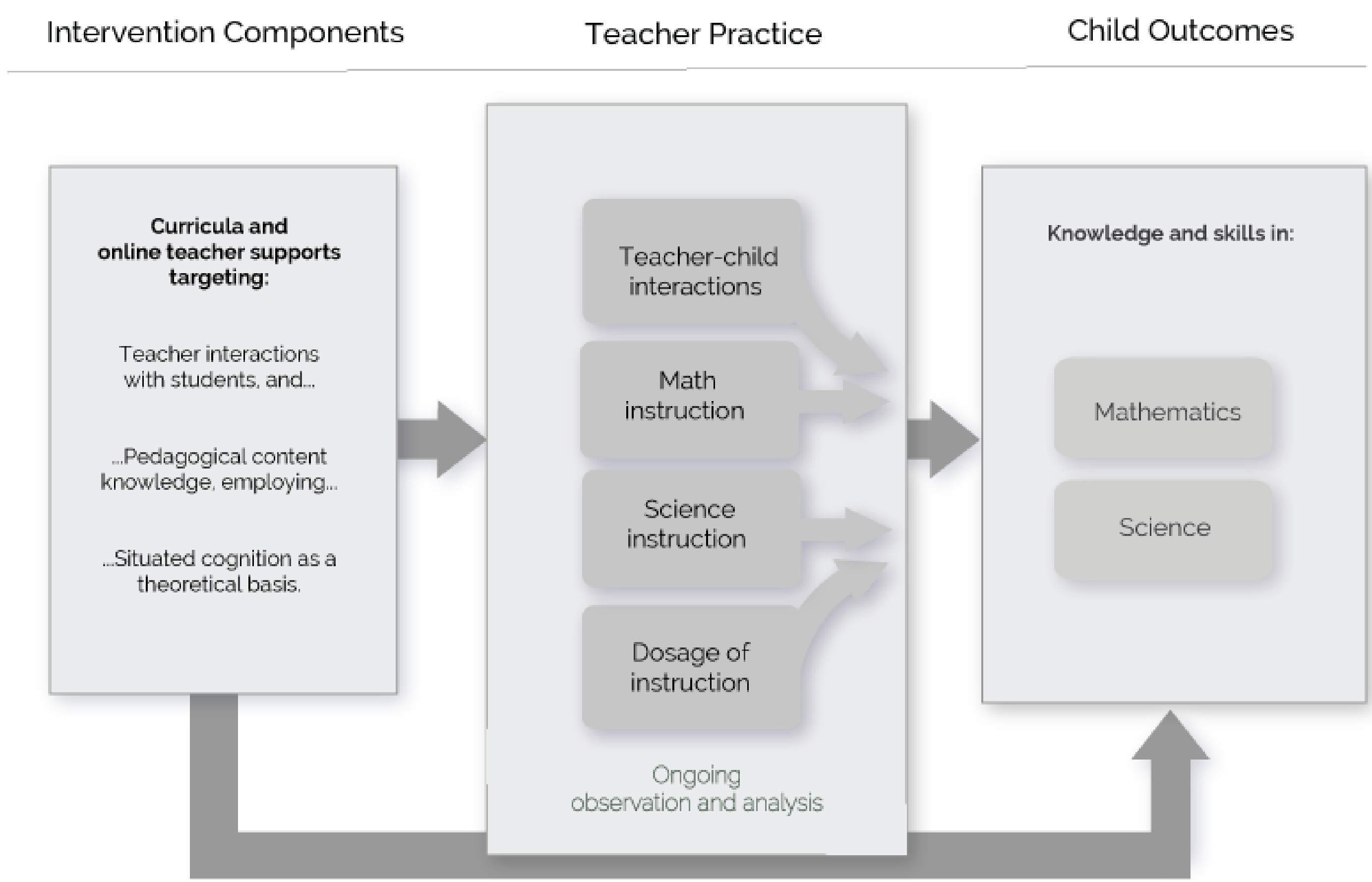

Figure 1. Logic Model

November-Science-W1-A1 | **Living vs. Non-Living I** | Whole Group

GET READY

Objectives	**Topic(s)**	**Use the Lingo**	
• Describe characteristics of living and non-living things • Classify things as living or non-living	Life Science Humans Animals Plants	• Living • Non-Living • Grow • Dormant	• Seed • Plant • Animal • Hard, smooth, etc. to describe seeds

Materials: (☆ Provided; ➡ Not Provided)

- ☆ *What's Alive?* by Kathleen Zoehfeld
- ☆ Ziploc bags (one per student)
- ☆ Seeds (two per student)
- ☆ Tape (optional)
- ➡ Paper towels (one per student)
- ➡ Water
- ➡ Stapler

Additional Preparation Required:

- Write student names on the Ziploc bags.
- Optional: In order to accelerate the plant growth in this activity, have children soak two paper towels with water and place the seeds between them. Let sit overnight and then proceed with this activity.

Teaching Tip: *Big Idea and How Kids Think*

Seeds are alive, but often dormant (sleeping) until they are given air, water and, in many cases, soil and warmth.

For preschoolers, it may be difficult to explain this in detail. To help them understand, explain that they themselves need food, but they are not constantly eating. But they cannot wait *too* long to eat.

The same goes for seeds and their needs – seeds can wait in a dormant (sleeping) state for some time but eventually will need water and air to grow.

ENGAGE

1. Science Chant.
2. Say: *Today we are going to learn what it means to be alive by reading this book. Then we will help something that is alive to grow!*
3. Show children the seeds. Tell the students that these seeds are alive but they need our help to grow. Say: *Today we'll learn what our seeds need to be alive and to grow!*

INVESTIGATE

4. Read *What's Alive?* by Kathleen Zoehfeld (up to p. 22, stop at "Now you can go exploring!").
   - Ask how students are similar to or different from other living things (e.g., a tree, flower, dog, etc.) and non-living things (e.g., a doll, chair, jar, etc.).
   - Ask students if specific things depicted in the book are living or non-living.
5. Discuss the characteristics of living things:
   - Explain to the students that all living things need food, water, and air and that all living things grow. Say: *Our seeds already have food from their mother plant. If we want to help our seeds, we must give them water and air.*
6. Investigate seeds.
   - Give each student a paper towel and Ziploc bag.
   - Ask students to dampen the paper towel. Explain that water in the towel will help the seed grow.
   - Pass out two seeds per student. Encourage students to examine and describe their seeds. (Optional: Provide hand lenses.)
   - Ask each student to place the dampened paper towel in the bag and then place the seeds in the bag on top of the towel (so they are easily visible).
   - Position the seeds in the middle of the bag and fix their position, by stapling the bag below the seeds (this allows the roots to grow downward).
   - Tape or staple the bags to a wall or board, so that students' names are visible.
   - Ask the students: *What do the seeds need in order to grow? Do our seeds have those things now? Do you think they will grow?*
   - During the next few days, ask students to observe if their seeds sprout. Ask students again if the seeds are alive and why. Allow time for an observation of the sprouted seeds using the hand lenses.

DISCUSS

7. Say: *Today we learned about what living things need and provided air and water to our seeds to help them grow!*
   - *Are you alive? Am I alive? Is the rug alive? Are seeds alive? What does it mean to be 'alive'? Why do you think that?*
   - *What did we give our seeds? What will happen? Why?*
   - *Can you see something in the room that is alive? How do you know? Why?*

   Discuss specific items in relation to criteria for being alive.

EXTEND

Point out living, non-living, and plant and animal materials throughout the day around the school.

- If possible, point out items that were specifically addressed in *What's Alive?* by Kathleen Zoehfeld.

MAKE IT WORK

For Students Requiring More Challenge	For Students Requiring More Support
• Expand by distinguishing between plants and animals. Ask students if plants and animals are both alive. *How do animals/plants get their food, air and water?*	• Provide a visual cue for each prompt (food, water, air, and grow). Examine living and non-living things by individually checking each characteristic using the visual cues.

Figure 2. Sample Activity for MTP-M/S Science

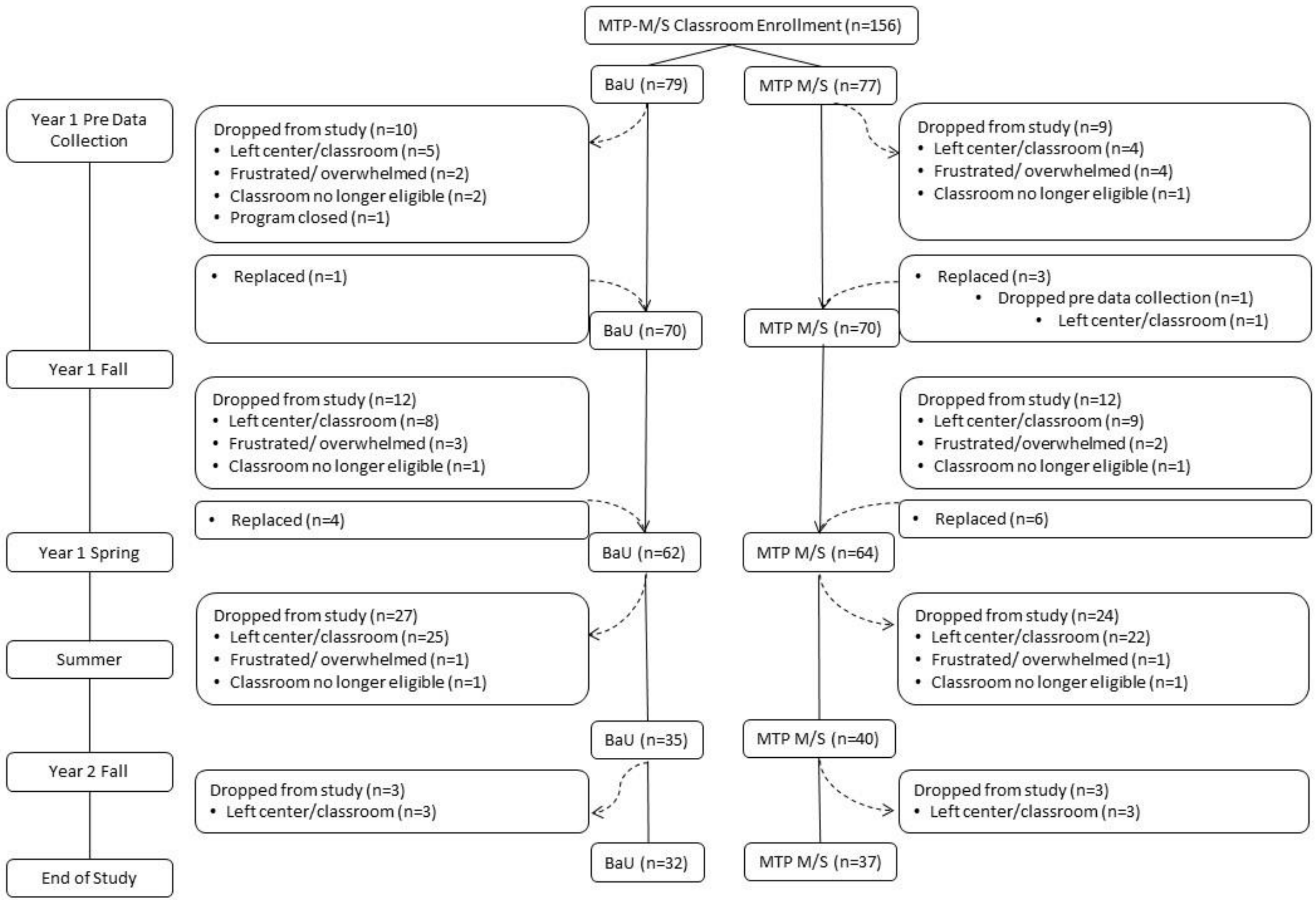

Figure 3. MTP Consort Flow Diagram

Appendix A: Tables that Show Results from Models that Include the Propensity Score as a Covariate

Table 10

Standardized Coefficients from a Sensitivity Analysis Relating Treatment to Classroom Outcomes Controlling for Covariates including Propensity Scores

Year 2 Outcome	*beta*	*SE{beta}*	*p*-value
Mathematics quality	.704	.082	.000
Science quality	.749	.081	.000
CLASS Emotional Support	-.081	.122	.504
CLASS Classroom Organization	.055	.131	.675
CLASS Instructional Support	.259	.119	.030
Proportion of time engaged in math	.263	.111	.018
Proportion of time engaged in science	.247	.111	.026

Note. Treatment was coded 1 = Treatment, 0 = Control, so positive coefficients represent more positive changes in the treatment condition than in the control condition.

Table 11

Standardized Coefficients from a Sensitivity Analysis Relating Treatment to Spring Child Outcomes Controlling for Fall Child Outcomes and Covariates including Propensity Scores

Year 2 Outcome	*beta*	*SE{beta}*	*p*-value
ARS Math	.264	.087	.002
NSPV	.116	.088	.188
MTP-STEAM	.044	.072	.540
ARS Science	.270	.093	.004
LENS	.092	.070	.190
PAS	.216	.089	.015